

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:14:12 ; Search time 25.7323 Seconds  
(without alignments)  
265.240 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	% Query		Match Length	DB	ID	Description
	Score	Match				
1	212	100.0	43	21	AAB18941	Peptide derived fr
2	212	100.0	84	21	AAB18942	Peptide derived fr
3	212	100.0	174	21	AAB18943	Peptide derived fr
4	212	100.0	186	21	AAB18944	Peptide derived fr
5	212	100.0	540	17	AAW07871	GDU (or Grb14), a
6	205	96.7	43	21	AAB18937	Peptide derived fr
7	205	96.7	84	21	AAB18938	Peptide derived fr
8	205	96.7	174	21	AAB18939	Peptide derived fr
9	205	96.7	186	21	AAB18940	Peptide derived fr
10	190	89.6	178	22	ABG02112	Novel human diagno
11	169	79.7	43	21	AAB18949	Peptide derived fr
12	169	79.7	82	21	AAB18950	Peptide derived fr
13	169	79.7	172	21	AAB18951	Peptide derived fr
14	169	79.7	184	21	AAB18952	Peptide derived fr
15	169	79.7	536	20	AAW83013	Human growth facto
16	169	79.7	594	22	AAB98060	Human SH2 and plec
17	169	79.7	723	22	ABG01373	Novel human diagno
18	162	76.4	43	21	AAB18957	Peptide derived fr
19	162	76.4	80	21	AAB18958	Peptide derived fr
20	162	76.4	170	21	AAB18959	Peptide derived fr
21	162	76.4	182	21	AAB18960	Peptide derived fr
22	162	76.4	329	23	ABP41924	Human ovarian anti
23	162	76.4	498	22	AAB93348	Human protein sequ
24	162	76.4	532	23	ABG96335	Human ovarian canc
25	161	75.9	43	21	AAB18945	Peptide derived fr
26	161	75.9	82	21	AAB18946	Peptide derived fr
27	161	75.9	172	21	AAB18947	Peptide derived fr
28	161	75.9	184	21	AAB18948	Peptide derived fr
29	161	75.9	326	16	AAR80162	GRB-10 central BLM
30	161	75.9	596	22	AAB98059	Mouse Meg1/Grb10 p
31	161	75.9	618	16	AAR80165	Mouse signal trans
32	161	75.9	621	16	AAR85785	Human GRB-10. Hom
33	159	75.0	43	21	AAB18953	Peptide derived fr
34	159	75.0	43	21	AAB18961	Peptide derived fr
35	159	75.0	80	21	AAB18954	Peptide derived fr
36	159	75.0	80	21	AAB18962	Peptide derived fr
37	159	75.0	170	21	AAB18955	Peptide derived fr
38	159	75.0	170	21	AAB18963	Peptide derived fr
39	159	75.0	182	21	AAB18956	Peptide derived fr
40	159	75.0	182	21	AAB18964	Peptide derived fr
41	159	75.0	334	16	AAR80167	Mouse signal trans
42	159	75.0	334	16	AAR80220	GRB-7 adaptor prot
43	159	75.0	335	16	AAR80161	GRB-7 central BLM
44	159	75.0	534	16	AAR80164	Mouse signal trans
45	159	75.0	535	16	AAR86900	Human GRB-7. Homo

# ALIGNMENTS

RESULT 1

AAB18941

ID AAB18941 standard; peptide; 43 AA.

XX

AC AAB18941;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH.SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 25; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 43 AA;

Query Match 100.0%; Score 212; DB 21; Length 43;

Best Local Similarity 100.0%; Pred. No. 9.2e-25;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

||||||||||||||||||||||||||||||||||||||||

Db 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

RESULT 2

AAB18942

ID AAB18942 standard; peptide; 84 AA.

XX

AC AAB18942;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity  
PT -

XX

PS Claim 2; Page 26; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 84 AA;

Query Match 100.0%; Score 212; DB 21; Length 84;

Best Local Similarity 100.0%; Pred. No. 2.3e-24;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43



Db

|||||  
13 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 55

RESULT 3.

AAB18943

ID AAB18943 standard; peptide; 174 AA.

XX

AC AAB18943;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 26; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 174 AA;

Query Match 100.0%; Score 212; DB 21; Length 174;

Best Local Similarity 100.0%; Pred. No. 6.3e-24;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||||||||||||||||||||||||||||||||||||||  
 Db 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

RESULT 4

AAB18944

ID AAB18944 standard; peptide; 186 AA.

XX

AC AAB18944;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity

PT -

XX

PS Claim 2; Page 27; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.

XX

SQ Sequence 186 AA;

Query Match 100.0%; Score 212; DB 21; Length 186;



CC altered expression of GDU may also contribute to abnormal cell  
CC proliferation, invasion and/or migration of cancer cells, GDU  
CC singnal transduction may provide a novel therapeutic target in  
CC human cancer. GDU is involved in downstream signalling initiated by  
CC platelet deriv. growth factor receptor (PDGFr), and may therefore  
CC provide a target in diseases or conditions in which PDGFr plays a  
CC regulatory role, e.g. wound healing, fibrotic conditions and  
CC atherosclerosis.

XX

SQ Sequence 540 AA;

Query Match 100.0%; Score 212; DB 17; Length 540;  
Best Local Similarity 100.0%; Pred. No. 3e-23;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
|||  
Db 367 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 409

#### RESULT 6

AAB18937

ID AAB18937 standard; peptide; 43 AA.

XX

AC AAB18937;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT -

XX

PS Claim 2; Page 23; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting

CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 43 AA;

Query Match 96.7%; Score 205; DB 21; Length 43;  
 Best Local Similarity 93.0%; Pred. No. 1.1e-23;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||:|||||||:||||:|||||||  
 Db 1 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 43

# RESULT 7

AAB18938

ID AAB18938 standard; peptide; 84 AA.

XX

AC AAB18938;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 23-24; 46pp; French.

XX  
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 84 AA;

Query Match 96.7%; Score 205; DB 21; Length 84;  
Best Local Similarity 93.0%; Pred. No. 2.7e-23;  
Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

||||:|||||||:||||:|||||||

Db 13 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 55

#### RESULT 8

AAB18939

ID AAB18939 standard; peptide; 174 AA.

XX

AC AAB18939;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT -

XX  
 PS Claim 2; Page 24; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 174 AA;

Query Match 96.7%; Score 205; DB 21; Length 174;  
 Best Local Similarity 93.0%; Pred. No. 7.4e-23;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||:|||||||:||||:|||||||  
 Db 1 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 43

# RESULT 9

AAB18940

ID AAB18940 standard; peptide; 186 AA.  
 XX  
 AC AAB18940;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX  
 OS Rattus sp.  
 XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in

PT treating insulin-associated diseases, particularly diabetes and obesity  
PT -  
XX  
PS Claim 2; Page 24-25; 46pp; French.  
XX  
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.  
XX  
SQ Sequence 186 AA;

Query Match 96.7%; Score 205; DB 21; Length 186;  
Best Local Similarity 93.0%; Pred. No. 8.1e-23;  
Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
|||:|||||||:|||||||:|||||||  
Db 13 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 55

#### RESULT 10

ABG02112

ID ABG02112 standard; Protein; 178 AA.

XX

AC ABG02112;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #2103.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.



DR N-PSDB; AAS66299.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX

PS Claim 20; SEQ ID No 32471; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 178 AA;

Query Match 89.6%; Score 190; DB 22; Length 178;  
Best Local Similarity 100.0%; Pred. No. 1.5e-20;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||  
Db 92 RSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 130

#### RESULT 11

AAB18949

ID AAB18949 standard; peptide; 43 AA.

XX

AC AAB18949;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Homo sapiens.

XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -  
 XX  
 PS Claim 2; Page 30; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 43 AA;

Query Match 79.7%; Score 169; DB 21; Length 43;  
 Best Local Similarity 76.7%; Pred. No. 3.4e-18;  
 Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |:||:||||||| ||||| || | |:||| ||||:  
 Db 1 PVRVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKR 43

# RESULT 12

AAB18950

ID AAB18950 standard; peptide; 82 AA.

XX

AC AAB18950;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX  
 OS Homo sapiens.  
 XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -  
 XX  
 PS Claim 2; Page 30; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 82 AA;

Query Match 79.7%; Score 169; DB 21; Length 82;  
 Best Local Similarity 76.7%; Pred. No. 8.2e-18;  
 Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |::|||:||||| ||||| || | :||| ||||:  
 Db 13 PVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKR 55

# RESULT 13

AAB18951

ID AAB18951 standard; peptide; 172 AA.

XX

AC AAB18951;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;









AC ABG01373;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #1364.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS65560.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 31732; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 723 AA;





XX

50

Oy

Db

AAB18958

ID

XX

AC

XX

DT

XX

DE

XX

KW

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

XX

PA

XX

PI

XX

DR

XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC





CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 182 AA;

Query Match 76.4%; Score 162; DB 21; Length 182;  
Best Local Similarity 74.4%; Pred. No. 2.9e-16;  
Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGIAWRKK 43  
|:| | |:|:||||||| | ||||| |||||:| | |||||  
Db 13 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKK 55

#### RESULT 22

ABP41924

ID ABP41924 standard; Protein; 329 AA.

XX

AC ABP41924;

XX

DT 22-AUG-2002 (first entry)

XX

DE Human ovarian antigen HODKM52, SEQ ID NO:3056.

XX

KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive.

XX

OS Homo sapiens.

XX

PN WO200200677-A1.

XX

PD 03-JAN-2002.

XX

PF 07-JUN-2001; 2001WO-US18569.

XX

PR 07-JUN-2000; 2000US-209467P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Birse CE, Rosen CA;

XX

DR WPI; 2002-147878/19.

DR N-PSDB; ABQ55001.

XX

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,



KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX

PA (HELI-) HELIX RES INST.

XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

XX

PS Claim 8; SEQ ID 12468; 2537pp + CD ROM; English.

XX

CC The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX

SQ Sequence 498 AA;

Query Match 76.4%; Score 162; DB 22; Length 498;  
Best Local Similarity 74.4%; Pred. No. 1.1e-15;

Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```
QY      1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43
      |:| |:|:||||| | |||| | ||||:| | ||||
Db     329 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRKK 371
```

RESULT 24

ABG96335

ID ABG96335 standard; Protein; 532 AA.

XX

AC ABG96335;

XX

DT 11-DEC-2002 (first entry)

XX

DE Human ovarian cancer marker M447.

XX

KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;  
KW central nervous system disorder; bacterial meningitis; viral meningitis;  
KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;  
KW brain herniation; inflammation; encephalitis; testicular disorder;  
KW nontuberculous granulomatous orchitis; connective tissue disorder;  
KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;  
KW histological type; carcinogenic; ovarian cancer marker.

XX

OS Homo sapiens.

XX

PN WO200271928-A2.

XX

PD 19-SEP-2002.

XX

PF 14-MAR-2002; 2002WO-US07826.

XX

PR 14-MAR-2001; 2001US-276025P.

PR 14-MAR-2001; 2001US-276026P.

PR 10-AUG-2001; 2001US-311732P.

PR 19-SEP-2001; 2001US-323580P.

PR 26-SEP-2001; 2001US-324967P.

PR 26-SEP-2001; 2001US-325102P.

PR 26-SEP-2001; 2001US-325149P.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;  
PI Meyers RE, Morrissey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;  
PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;

XX

DR WPI; 2002-723277/78.

DR N-PSDB; ABS76431.

XX

PT Assessing whether a patient is afflicted with ovarian cancer, useful in  
PT assessing the stage or progression of the disease, comprises comparing  
PT the expression level of a cancer marker in a sample from a patient and  
PT from a non cancer patient -

XX

PS Disclosure; Page 245-246; 481pp; English.

XX



CC The present invention relates to a new method for assessing whether a  
CC patient is afflicted with ovarian cancer. The method involves comparing  
CC the expression level of a marker in a patient sample and the normal level  
CC of expression of the marker in a control non-ovarian cancer sample, where  
CC the marker is selected from 363 cancer markers described in the  
CC specification. The method of the invention is useful in diagnosing or  
CC characterising cancer, in detecting the presence of cancer as early as  
CC possible, and the recurrence of ovarian cancer. The method may also be of  
CC particular use with patients having an enhanced risk of developing  
CC ovarian cancer (e.g. patients having a familial history of ovarian  
CC cancer). The cancer markers may be used in the management and treatment  
CC of e.g. brain and central nervous system disorders (e.g. bacterial and  
CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain  
CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),  
CC inflammations (e.g. bacterial or viral meningitis or encephalitis),  
CC testicular disorders (e.g. nontuberculous granulomatous orchitis),  
CC connective tissue disorders, or heart disorders (e.g. ischaemic heart  
CC disease or atherosclerosis). The compositions and methods may also be  
CC used in assessing the histological type of neoplasm associated with  
CC ovarian cancer, monitoring the progression of ovarian cancer,  
CC determining whether ovarian cancer has metastasized or is likely to  
CC metastasize, selecting a composition for inhibiting ovarian cancer,  
CC assessing the ovarian carcinogenic potential of a compound, or  
CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The  
CC present amino acid sequence represents one of the ovarian cancer markers  
CC described in the invention.

XX

SQ Sequence 532 AA;

Query Match 76.4%; Score 162; DB 23; Length 532;  
Best Local Similarity 74.4%; Pred. No. 1.3e-15;  
Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
|:| | |:|:||||||| ||||| |||||:| | |||||  
Db 363 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRKK 405

#### RESULT 25

AAB18945

ID AAB18945 standard; peptide; 43 AA.

XX

AC AAB18945;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus muris.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -  
 XX  
 PS Claim 2; Page 27-28; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 43 AA;

Query Match 75.9%; Score 161; DB 21; Length 43;  
 Best Local Similarity 78.0%; Pred. No. 5.6e-17;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| || | :||| |||  
 Db 1 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 41

# RESULT 26

AAB18946

ID AAB18946 standard; peptide; 82 AA.

XX

AC AAB18946;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus muris.

XX

PN WO200055634-A1.

XX  
PD 21-SEP-2000.  
XX  
PF 14-MAR-2000; 2000WO-FR00613.  
XX  
PR 15-MAR-1999; 99FR-0003159.  
XX  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
XX  
DR WPI; 2000-587566/55.  
XX  
PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity  
PT -  
XX  
PS Claim 2; Page 28; 46pp; French.  
XX  
CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.  
XX  
SQ Sequence 82 AA;

Query Match 75.9%; Score 161; DB 21; Length 82;  
Best Local Similarity 78.0%; Pred. No. 1.4e-16;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
| | | | : | | | | | | | | | | | | | | : | | | | : | | | | |  
Db 13 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 53

# RESULT 27

AAB18947

ID AAB18947 standard; peptide; 172 AA.

XX

AC AAB18947;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus muris.

XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -  
 XX  
 PS Claim 2; Page 28-29; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 172 AA;

Query Match 75.9%; Score 161; DB 21; Length 172;  
 Best Local Similarity 78.0%; Pred. No. 3.8e-16;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| | | |:| | |  
 Db 1 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 41

# RESULT 28

AAB18948

ID AAB18948 standard; peptide; 184 AA.

XX

AC AAB18948;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX  
 OS Mus muris.  
 XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -  
 XX  
 PS Claim 2; Page 29; 46pp; French.  
 XX  
 CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 184 AA;

Query Match 75.9%; Score 161; DB 21; Length 184;  
 Best Local Similarity 78.0%; Pred. No. 4.1e-16;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| || | |:| || ||  
 Db 13 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 53

# RESULT 29

AAR80162

ID AAR80162 standard; peptide; 326 AA.

XX

AC AAR80162;

XX

DT 22-APR-1996 (first entry)

XX

DE GRB-10 central BLM domain.

XX

KW Signal transduction protein; growth factor receptor bound; BLM domain;

KW pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;  
 KW abnormal cell development; cell movement; breast cancer; atherosclerosis.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO9525166-A1.  
 XX  
 PD 21-SEP-1995.  
 XX  
 PF 13-MAR-1995; 95WO-US03452.  
 XX  
 PR 08-JUN-1994; 94US-0255785.  
 PR 14-MAR-1994; 94US-0212234.  
 XX  
 PA (UYNY-) UNIV NEW YORK MEDICAL CENT.  
 XX  
 PI Ladbury JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;  
 XX  
 DR WPI; 1995-336971/43.  
 XX  
 PT Treating diseases involving abnormal signal transduction e.g. cancer  
 PT and psoriasis - by modulating interaction between e.g. epidermal  
 PT growth factor receptor and its ligand, also diagnosis and screening  
 PT of modulators  
 XX  
 PS Disclosure; Fig 2; 102pp; English.  
 XX  
 CC The amino acid sequence of the central domain of the signal transduction  
 CC protein, growth factor receptor bound (GRB)-10 protein. The protein  
 CC contains a central BLM domain and within this domain a pleckstrin domain.  
 CC The central domain is flanked by a proline-rich and an SH2 domain  
 CC indicating that the protein is involved in signal transduction. The SH2  
 CC domain has been shown to bind to the HER2 receptor protein. The protein  
 CC can be used to screen for cpds. which can promote or interrupt  
 CC interaction of proteins involved in signal transduction, esp. in neuronal  
 CC diseases, diseases involved with abnormal cell development and defective  
 CC cell movement, breast cancer, atherosclerosis, etc.  
 XX  
 SQ Sequence 326 AA;

Query Match 75.9%; Score 161; DB 16; Length 326;  
 Best Local Similarity 78.0%; Pred. No. 9.1e-16;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| || | :||| |||  
 Db 262 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 302

# RESULT 30

AAB98059

ID AAB98059 standard; Protein; 596 AA.

XX

AC AAB98059;

XX

DT 15-AUG-2001 (first entry)

XX

DE Mouse Meg1/Grb10 protein sequence SEQ ID NO:2.  
 XX  
 KW Mouse; Meg1/Grb10; diabetes; transgene; transgenic animal;  
 KW insulin signal transduction inhibition.  
 XX  
 OS Mus sp.  
 XX  
 PN WO200128321-A1.  
 XX  
 PD 26-APR-2001.  
 XX  
 PF 18-AUG-2000; 2000WO-JP05546.  
 XX  
 PR 20-OCT-1999; 99JP-0298273.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Ishino F, Miyoshi N, Ishino T, Yokoyama M, Wakana S;  
 XX  
 DR WPI; 2001-300253/31.  
 DR N-PSDB; AAH21792, AAH21793.  
 XX  
 PT Transgenic non-human mammal with Meg1/Grb10 or human GRB 10 gene useful  
 PT as a model for onset of diabetes and for screening new diabetes  
 PT treatments -  
 XX  
 PS Claim 2; Page 30-31; 50pp; Japanese.  
 XX  
 CC The present invention describes a transgenic non-human mammal containing  
 CC the Meg1/Grb10 gene. Also described are: (1) a transgenic non human  
 CC mammal with human GRB10 gene; (2) a method for producing a transgenic  
 CC mouse; (3) method (M1) for screening for drugs for treating diabetes;  
 CC and (4) drugs found using (M1). The transgenic non-human mammal is  
 CC useful for screening for new drugs to treat diabetes. The transgenic  
 CC animals are models for the onset of diabetes, and may be useful in  
 CC discovering the mechanism for the onset of diabetes caused by inhibition  
 CC of insulin signal transduction, and for developing new treatments. The  
 CC present sequence represents a specifically claimed mouse Meg1/Grb10  
 CC protein sequence from the present invention.  
 XX  
 SQ Sequence 596 AA;

Query Match 75.9%; Score 161; DB 22; Length 596;  
 Best Local Similarity 78.0%; Pred. No. 2.1e-15;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| | | |:| | |  
 Db 425 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 465

RESULT 31  
 AAR80165  
 ID AAR80165 standard; peptide; 618 AA.  
 XX  
 AC AAR80165;  
 XX

DT 22-APR-1996 (first entry)  
 XX  
 DE Mouse signal transduction protein GRB-10.  
 XX  
 KW Signal transduction protein; growth factor receptor bound; BLM domain;  
 KW pleckstrin domain; SH2 domain; HER2 receptor; mouse; neuronal disease;  
 KW abnormal cell development; cell movement; breast cancer; atherosclerosis.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO9525166-A1.  
 XX  
 PD 21-SEP-1995.  
 XX  
 PF 13-MAR-1995; 95WO-US03452.  
 XX  
 PR 08-JUN-1994; 94US-0255785.  
 PR 14-MAR-1994; 94US-0212234.  
 XX  
 PA (UYNY-) UNIV NEW YORK MEDICAL CENT.  
 XX  
 PI Ladbury JE, Lax I, Lemmon MA, Margolis BL, Schlessinger J;  
 XX  
 DR WPI; 1995-336971/43.  
 XX  
 PT Treating diseases involving abnormal signal transduction e.g. cancer  
 PT and psoriasis - by modulating interaction between e.g. epidermal  
 PT growth factor receptor and its ligand, also diagnosis and screening  
 PT of modulators  
 XX  
 PS Disclosure; Fig 3; 102pp; English.  
 XX  
 CC The amino acid sequence of the signal transduction protein, growth  
 CC factor receptor bound (GRB)-10 protein. This sequence covers from amino  
 CC acids 4-621 of the full length protein. The protein contains a central  
 CC BLM domain and within this domain a pleckstrin domain (AAR80162). The  
 CC central domain is flanked by a proline-rich and an SH2 domain indicating  
 CC that the protein is involved in signal transduction. The SH2 domain has  
 CC been shown to bind to the HER2 receptor protein. The protein can be used  
 CC to screen for cpds. which can promote or interrupt interaction of  
 CC proteins involved in signal transduction, esp. in neuronal diseases,  
 CC diseases involved with abnormal cell development and defective cell  
 CC movement, breast cancer, atherosclerosis, etc.  
 XX  
 SQ Sequence 618 AA;

Query Match 75.9%; Score 161; DB 16; Length 618;  
 Best Local Similarity 78.0%; Pred. No. 2.2e-15;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| || | |:| || ||  
 Db 447 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 487

RESULT 32  
 AAR85785



ID AAR85785 standard; Protein; 621 AA.  
 XX  
 AC AAR85785;  
 XX  
 DT 16-MAY-1996 (first entry)  
 XX  
 DE Human GRB-10.  
 XX  
 KW GRB-10; growth factor receptor bound; tyrosine kinase; regulation;  
 KW cell growth; cellular metabolism; screening; signal transduction;  
 KW cancer; diabetes; CORT technique; cloning of receptor targets.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9524426-A1.  
 XX  
 PD 14-SEP-1995.  
 XX  
 PF 13-MAR-1995; 95WO-US03385.  
 XX  
 PR 11-MAR-1994; 94US-0208887.  
 XX  
 PA (UYN Y ) UNIV NEW YORK STATE.  
 XX  
 PI Margolis BL, Schlessinger J, Skolnik EY;  
 XX  
 DR WPI; 1995-328235/42.  
 DR N-PSDB; AAT03197.  
 XX  
 PT DNA encoding tyrosine kinase-binding proteins - used to screen  
 PT agents capable of modulating cell growth or cellular metabolism  
 XX  
 PS Claim 1; Fig 38; 215pp; English.  
 XX  
 CC Using a new cloning technique, CORT (cloning of receptor targets)  
 CC several new tyrosine kinase (TK) binding proteins were isolated. Growth  
 CC factor receptor bound proteins GRB-1, GRB-2, GRB-3, GRB-4, GRB-7 and  
 CC GRB-10 were isolated using this method. This sequence represents GRB-10.  
 CC The proteins bind to a tyrosine-phosphorylated domain of a eukaryotic  
 CC TK. GRB proteins can be used for screening agents which are capable  
 CC of modulating cell growth that occurs via signal transduction through  
 CC TKs. Such agents can be used to prevent or inhibit cell growth or to  
 CC counteract tumour development. GRB proteins are also useful for  
 CC identifying susceptibility to diseases associated with alterations in  
 CC cellular metabolism mediated by TK pathways e.g. cancer and diabetes.  
 XX  
 SQ Sequence 621 AA;  
  
 Query Match 75.9%; Score 161; DB 16; Length 621;  
 Best Local Similarity 78.0%; Pred. No. 2.2e-15;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
  
 QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| | | | :||| |||  
 Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 490

RESULT 33

AAB18953

ID AAB18953 standard; peptide; 43 AA.

XX

AC    AAB18953;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS     Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

P<sup>T</sup>

XX

PS Claim 2; Page 32; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.

XX

SQ Sequence 43 AA;

Query Match 75.0%; Score 159; DB 21; Length 43;

Best Local Similarity 69.8%; Pred. No. 1.1e-16;

Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

| : | | : | : | : | ||||| | | | | : | | | | | | : | | | | |

Db

1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAOAWRKK 43

RESULT 34

AAB18961

ID AAB18961 standard; peptide; 43 AA.

XX

AC AAB18961;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus muris.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 36; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 43 AA;

Query Match 75.0%; Score 159; DB 21; Length 43;

Best Local Similarity 69.8%; Pred. No. 1.1e-16;

Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

Db

1 P L R S V S D N T L V A M D F S G H A G R V I D N P R E A L S A A M E E A Q A W R K K 43

RESULT 35

AAB18954

ID AAB18954 standard; peptide; 80 AA.

XX

AC AAB18954;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Rattus sp.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 32; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX

SQ Sequence 80 AA;

Query Match 75.0%; Score 159; DB 21; Length 80;

Best Local Similarity 69.8%; Pred. No. 2.7e-16;

Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY            1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK    43  
| : | : | : | : | | | | | |       | | : | | | | | : | |    | | | |  
Db           13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK    55

RESULT 36

AAB18962

ID AAB18962 standard; peptide; 80 AA.

XX

AC AAB18962;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus muris.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 37; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.

CC

CC greater in presence of SH2 (which by itself is inactive). Agents that

CC

stimulate or inhibit tyrosine kinase activity of the receptor. The peptides are used for screening molecules for ability to treat diseases in which insulin is implicated. The peptides are used to identify agents that are potentially useful for treating insulin-associated diseases, particularly diabetes and obesity but also polycystic ovarian syndrome and syndrome X.

XX

SQ Sequence 80 AA;

## Query Match

75.0%; Score 159; DB 21; Length 80;



Query Match 75.0%; Score 159; DB 21; Length 170;  
Best Local Similarity 69.8%; Pred. No. 7.5e-16;  
Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY           1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
| : | : | : | ||||| |     || : | | || | | : | |     ||||  
Db          1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 43

RESULT 38

AAB18963

ID AAB18963 standard; peptide; 170 AA.

XX

AC AAB18963;

XX

DT 08-FEB-2001 (first entry)

XX

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;

KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX

OS Mus muris.

XX

PN WO200055634-A1.

XX

PD 21-SEP-2000.

XX

PF 14-MAR-2000; 2000WO-FR00613.

XX

PR 15-MAR-1999; 99FR-0003159.

XX

PA (CNRS ) CNRS CENT NAT RECH SCI .

XX

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX

DR WPI; 2000-587566/55.

XX

PT Fragments of Grb family proteins to identify compounds are useful in

PT treating insulin-associated diseases, particularly diabetes and obesity

PT

XX

PS Claim 2; Page 37-38; 46pp; French.

XX

CC B18937-64 represent the PIR (phosphorylated insulin receptor interacting

CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.

CC. PIR is the actual binding region but its effect is about 10 times

CC greater in presence of SH2 (which by itself is inactive). Agents that

CC affect binding between the peptides and the insulin receptor can

CC stimulate or inhibit tyrosine kinase activity of the receptor. The

CC peptides are used for screening molecules for ability to treat diseases

CC in which insulin is implicated. The peptides are used to identify agents

CC that are potentially useful for treating insulin-associated diseases.

CC particularly diabetes and obesity but also polycystic ovarian syndrome

CC and syndrome X.







XX

2

Best Local Similarity 69.8%; Pred. No. 8.2e-16;

[illegible]

PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK

Job time : 26.7323 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:18:37 ; Search time 10.1575 Seconds  
(without alignments)  
179.116 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	212	100.0	540	4	US-08-945-771-2	Sequence 2, Appli
2	169	79.7	536	2	US-08-890-094-2	Sequence 2, Appli
3	169	79.7	548	2	US-08-890-094-18	Sequence 18, Appl
4	161	75.9	326	3	US-08-866-381A-2	Sequence 2, Appli
5	161	75.9	326	4	US-09-280-598-52	Sequence 52, Appl
6	161	75.9	618	3	US-08-866-381A-6	Sequence 6, Appli
7	161	75.9	621	1	US-08-208-887A-49	Sequence 49, Appl
8	161	75.9	621	4	US-09-280-598-18	Sequence 18, Appl
9	161	75.9	621	4	US-08-945-771-4	Sequence 4, Appli
10	159	75.0	334	3	US-08-472-595-9	Sequence 9, Appli
11	159	75.0	334	3	US-08-207-575A-9	Sequence 9, Appli

12	159	75.0	335	3	US-08-866-381A-1	Sequence 1, Appli
13	159	75.0	335	4	US-09-280-598-51	Sequence 51, Appl
14	159	75.0	534	3	US-08-866-381A-5	Sequence 5, Appli
15	159	75.0	535	1	US-07-906-349A-10	Sequence 10, Appl
16	159	75.0	535	1	US-08-167-035-10	Sequence 10, Appl
17	159	75.0	535	1	US-08-208-887A-10	Sequence 10, Appl
18	159	75.0	535	2	US-08-539-005-10	Sequence 10, Appl
19	159	75.0	535	4	US-09-280-598-10	Sequence 10, Appl
20	159	75.0	535	4	US-08-945-771-3	Sequence 3, Appli
21	59.5	28.1	1346	3	US-09-320-878-4	Sequence 4, Appli
22	59.5	28.1	1346	4	US-09-141-908-5	Sequence 5, Appli
23	59.5	28.1	1346	4	US-09-657-440-4	Sequence 4, Appli
24	58	27.4	1024	4	US-09-562-737-85	Sequence 85, Appl
25	56.5	26.7	1346	3	US-09-105-537-37	Sequence 37, Appl
26	56.5	26.7	11877	3	US-09-105-537-6	Sequence 6, Appli
27	55	25.9	480	4	US-09-107-532A-6160	Sequence 6160, Ap
28	55	25.9	823	4	US-09-252-991A-24768	Sequence 24768, A
29	54	25.5	315	4	US-09-328-352-6585	Sequence 6585, Ap
30	53	25.0	329	4	US-09-071-035-368	Sequence 368, App
31	53	25.0	357	4	US-09-071-035-366	Sequence 366, App
32	52.5	24.8	452	4	US-09-252-991A-31873	Sequence 31873, A
33	51	24.1	334	6	5290690-10	Patent No. 5290690
34	51	24.1	335	6	5290690-9	Patent No. 5290690
35	49	23.1	310	4	US-09-598-747-27	Sequence 27, Appl
36	49	23.1	399	4	US-09-252-991A-17604	Sequence 17604, A
37	49	23.1	443	4	US-09-328-352-6943	Sequence 6943, Ap
38	48	22.6	133	4	US-09-328-352-4244	Sequence 4244, Ap
39	48	22.6	295	2	US-08-454-267-7	Sequence 7, Appli
40	48	22.6	295	2	US-08-941-319-7	Sequence 7, Appli
41	48	22.6	295	3	US-09-035-098-7	Sequence 7, Appli
42	48	22.6	311	4	US-08-818-581B-6	Sequence 6, Appli
43	48	22.6	789	4	US-09-002-285-80	Sequence 80, Appl
44	48	22.6	789	4	US-09-589-477-80	Sequence 80, Appl
45	47.5	22.4	411	4	US-09-252-991A-23375	Sequence 23375, A

#### ALIGNMENTS

##### RESULT 1

US-08-945-771-2

; Sequence 2, Application US/08945771

; Patent No. 6465623

; GENERAL INFORMATION:

; APPLICANT: Daly, Roger J

; APPLICANT: Sutherland, Robert L

; TITLE OF INVENTION: GDU, A novel signalling protein

; FILE REFERENCE: 273402001700

; CURRENT APPLICATION NUMBER: US/08/945,771

; CURRENT FILING DATE: 1998-04-22

; EARLIER APPLICATION NUMBER: PCT/US96/00258

; EARLIER FILING DATE: 1996-MAY-02

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 540

; TYPE: PRT

; ORGANISM: Homo sapiens  
US-08-945-771-2

Query Match 100.0%; Score 212; DB 4; Length 540;  
Best Local Similarity 100.0%; Pred. No. 3.2e-24;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |||  
 Db 367 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 409

## RESULT 2

US-08-890-094-2

; Sequence 2, Application US/08890094

; Patent No. 5840536

GENERAL INFORMATION:

APPLICANT: SmithKline Beecham Corporation and Harvard University

TITLE OF INVENTION: GROWTH FACTOR RECEPTOR-BINDING INSULIN RECEPTOR

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

; CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

```
; SOFTWARE: FastSEQ Version 1.5
```

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/890,094

FILING DATE: 09-JULY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/022,703

; FILING DATE:

ATTORNEY/AGENT INFORMATION:

; NAME: Baumeister, Kirk

REGISTRATION NUMBER: 33,833

REFERENCE/DOCKET NUMBER: P50508P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5096

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 536 amino acids

; TYPE: amino acid

```
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear

```

; MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

```

; FRAGMENT TYPE: N-terminal

```





; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; OTHER INFORMATION: BLM domain of GRB-10  
US-08-866-381A-2

Query Match 75.9%; Score 161; DB 3; Length 326;  
Best Local Similarity 78.0%; Pred. No. 1.2e-16;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| | | |:| | |  
Db 262 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 302

RESULT 5

US-09-280-598-52

; Sequence 52, Application US/09280598

; Patent No. 6391584

; GENERAL INFORMATION:

; APPLICANT: Schlessinger, Joseph

; APPLICANT: Skolnik, Edward Y.

; APPLICANT: Margolis, Benjamin L.

; APPLICANT: App, Harold

; TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR

; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE

; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS

; NUMBER OF SEQUENCES: 58

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/280,598

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/252,820

; FILING DATE: 02-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Coruzzi, Laura A.

; REGISTRATION NUMBER: 30,742

; REFERENCE/DOCKET NUMBER: 7683-067

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE



; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 326 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-280-598-52

Query Match 75.9%; Score 161; DB 4; Length 326;  
Best Local Similarity 78.0%; Pred. No. 1.2e-16;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| | | |:| |  
Db 262 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 302

RESULT 6

US-08-866-381A-6

; Sequence 6, Application US/08866381A

; Patent No. 6045797

; GENERAL INFORMATION:

; APPLICANT: Ben Lewis Margolis

; APPLICANT: Joseph Schlessinger

; TITLE OF INVENTION: METHODS FOR TREATMENT OR DIAGNOSIS

; TITLE OF INVENTION: OF DISEASES OR CONDITIONS ASSOCIATED

; TITLE OF INVENTION: WITH A BLM DOMAIN

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; STREET: Suite 4700

; CITY: Los Angeles

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071-2066

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: FastSEQ for Windows 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/866,381A

; FILING DATE: May 30, 1997

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/212,234

; FILING DATE: March 14, 1994

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard J.

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 226/043

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 618 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FEATURE:  
 ; OTHER INFORMATION: GRB-10  
 US-08-866-381A-6

Query Match 75.9%; Score 161; DB 3; Length 618;  
 Best Local Similarity 78.0%; Pred. No. 2.9e-16;  
 Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||||:||||||| |||:| | | | :||| |||  
 Db 447 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 487

RESULT 7

US-08-208-887A-49  
 ; Sequence 49, Application US/08208887A  
 ; Patent No. 5677421  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schlessinger, Joseph  
 ; APPLICANT: Skolnick, Edward Y.  
 ; APPLICANT: Margolis, Benjamin L.  
 ; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
 ; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
 ; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: 10036-2711  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/208,887A  
 ; FILING DATE: 11-MAR-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7683-063  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 621 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-208-887A-49

Query Match 75.9%; Score 161; DB 1; Length 621;  
Best Local Similarity 78.0%; Pred. No. 3e-16;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| | | | :||| |||  
Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 490

RESULT 8

US-09-280-598-18

; Sequence 18, Application US/09280598  
; Patent No. 6391584  
; GENERAL INFORMATION:  
; APPLICANT: Schlessinger, Joseph  
; APPLICANT: Skolnik, Edward Y.  
; APPLICANT: Margolis, Benjamin L.  
; APPLICANT: App, Harold  
; TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR  
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/280,598  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/252,820  
; FILING DATE: 02-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7683-067  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 621 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-09-280-598-18

Query Match 75.9%; Score 161; DB 4; Length 621;  
Best Local Similarity 78.0%; Pred. No. 3e-16;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| || | |:||| |||  
Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 490

#### RESULT 9

US-08-945-771-4  
; Sequence 4, Application US/08945771  
; Patent No. 6465623  
; GENERAL INFORMATION:  
; APPLICANT: Daly, Roger J  
; APPLICANT: Sutherland, Robert L  
; TITLE OF INVENTION: GDU, A novel signalling protein  
; FILE REFERENCE: 273402001700  
; CURRENT APPLICATION NUMBER: US/08/945,771  
; CURRENT FILING DATE: 1998-04-22  
; EARLIER APPLICATION NUMBER: PCT/US96/00258  
; EARLIER FILING DATE: 1996-MAY-02  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 621  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-08-945-771-4

Query Match 75.9%; Score 161; DB 4; Length 621;  
Best Local Similarity 78.0%; Pred. No. 3e-16;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| || | |:||| |||  
Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 490

#### RESULT 10

US-08-472-595-9  
; Sequence 9, Application US/08472595  
; Patent No. 6001583  
; GENERAL INFORMATION:  
; APPLICANT: Margolis, Benjamin L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATMENT  
; TITLE OF INVENTION: OF BREAST CANCER  
; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/472,595  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7683-103  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 334 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 US-08-472-595-9

Query Match 75.0%; Score 159; DB 3; Length 334;  
 Best Local Similarity 69.8%; Pred. No. 2.6e-16;  
 Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |::|:|:|:||||||| |||:| | ||| |:| | |||||  
 Db 272 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 314

# RESULT 11

US-08-207-575A-9

; Sequence 9, Application US/08207575A  
 ; Patent No. 6037134

; GENERAL INFORMATION:  
 ; APPLICANT: Margolis, Benjamin L.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATMENT  
 ; TITLE OF INVENTION: OF BREAST CANCER  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.























```

; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-141-908-5

```

```

Query Match          28.1%; Score 59.5; DB 4; Length 1346;
Best Local Similarity 34.6%; Pred. No. 4.2;
Matches 18; Conservative 9; Mismatches 14; Indels 11; Gaps 2;

```

```

Qy      1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41
      |:| | :|| |:|| : :| | |:|| ||: : : || |
Db      972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFHPTPVALAERISDELAER 1023

```

# RESULT 23

US-09-657-440-4

```

; Sequence 4, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae

```



US-09-657-440-4

Query Match 28.1%; Score 59.5; DB 4; Length 1346;  
Best Local Similarity 34.6%; Pred. No. 4.2;  
Matches 18; Conservative 9; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : | | |:| | |: : : || |  
Db 972 PLREIGFDSLTAVD FRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023

RESULT 24

US-09-562-737-85

; Sequence 85, Application US/09562737  
; Patent No. 6428967  
; GENERAL INFORMATION:  
; APPLICANT: Herz, Joachim  
; APPLICANT: Gotthardt, Michael  
; TITLE OF INVENTION: LDL Receptor Signaling Pathways  
; FILE REFERENCE: UTSW0708  
; CURRENT APPLICATION NUMBER: US/09/562,737  
; CURRENT FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 85  
; LENGTH: 1024  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Sequence

US-09-562-737-85

Query Match 27.4%; Score 58; DB 4; Length 1024;  
Best Local Similarity 27.9%; Pred. No. 4.9;  
Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 1;

Qy 3 RSISENSLVAMDFSGQ--KSRVIENPTEALSVAVEEGLAWRKK 43  
::: || : |||| ::|: :|| :: : || :  
Db 460 QAVAANSAASRDFSGQGGLGELLESRSSEASKLSSKTAKWRNR 502

RESULT 25

US-09-105-537-37

; Sequence 37, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 37  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-37

Query Match 26.7%; Score 56.5; DB 3; Length 1346;  
Best Local Similarity 32.7%; Pred. No. 12;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : :| | :|| | |: : : || |  
Db 972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVVFQHPTPVALAERISDELAER 1023

RESULT 26

US-09-105-537-6

; Sequence 6, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 11877  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-6

Query Match 26.7%; Score 56.5; DB 3; Length 11877;  
Best Local Similarity 32.7%; Pred. No. 2.4e+02;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : :| | :|| | |: : : || |  
Db 11222 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVVFQHPTPVALAERISDELAER 11273

RESULT 27

US-09-107-532A-6160

; Sequence 6160, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

/ STREET: 100 Beaver Street  
 / CITY: Waltham  
 / STATE: Massachusetts  
 / COUNTRY: USA  
 / ZIP: 02354  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: CD-ROM ISO9660  
 / COMPUTER: PC  
 / OPERATING SYSTEM: <Unknown>  
 / SOFTWARE: ASCII  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/09/107,532A  
 / FILING DATE: 30-Jun-1998  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: 60/085,598  
 / FILING DATE: 14 May 1998  
 / APPLICATION NUMBER: 60/051571  
 / FILING DATE: July 2, 1997  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Ariniello, Pamela Deneke  
 / REGISTRATION NUMBER: 40,489  
 / REFERENCE/DOCKET NUMBER: GTC-012  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (781)893-5007  
 / TELEFAX: (781)893-8277  
 / INFORMATION FOR SEQ ID NO: 6160:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 480 amino acids  
 / TYPE: amino acid  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / HYPOTHETICAL: YES  
 / ORIGINAL SOURCE:  
 / ORGANISM: Enterococcus faecium  
 / FEATURE:  
 / NAME/KEY: misc\_feature  
 / LOCATION: (B) LOCATION 1...480  
 / SEQUENCE DESCRIPTION: SEQ ID NO: 6160:  
 US-09-107-532A-6160

Query Match 25.9%; Score 55; DB 4; Length 480;  
 Best Local Similarity 41.4%; Pred. No. 5;  
 Matches 12; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 10 LVAMDFSGQKSRVIENPTEALSVAVEEGL 38  
 || : | : : : : || || ||||  
 Db 301 LVCLGVIGEIASWVTSPSKALHVAAEEGL 329

# RESULT 28

US-09-252-991A-24768  
 ; Sequence 24768, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 PSEUDOMONAS

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24768
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24768

```

```

Query Match          25.9%; Score 55; DB 4; Length 823;
Best Local Similarity 32.5%; Pred. No. 10;
Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

```

```

QY      1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40
      | | : | | : | : | : | : | | | |
Db      522 PNRHGGRENLEAIDFLHHLNQVVAETPGALVIAEESTAW 561

```

# RESULT 29

```

US-09-328-352-6585
; Sequence 6585, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6585
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6585

```

```

Query Match          25.5%; Score 54; DB 4; Length 315;
Best Local Similarity 33.3%; Pred. No. 4;
Matches 13; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

```

```

QY      2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40
      : | | | : | | | | : | | | : |
Db      20 LTDIRLNPRTARNHSRSMKMSYENRWETVDVKVEDGIAW 58

```

# RESULT 30

```

US-09-071-035-368
; Sequence 368, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:

```

; APPLICANT: Gil H. Choi  
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
 ; NUMBER OF SEQUENCES: 496  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/071,035  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: A. Anders Brookes  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB369P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 368:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 329 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-071-035-368

Query Match 25.0%; Score 53; DB 4; Length 329;  
 Best Local Similarity 36.1%; Pred. No. 6;  
 Matches 13; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

Qy 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEG 37  
 | :| | ||: |||: : ::|| | ||  
 Db 195 MYMANEESAVAVTFSGEAAEMLENNEHLHYVIPSEG 230

RESULT 31

US-09-071-035-366

; Sequence 366, Application US/09071035  
 ; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi  
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
 ; NUMBER OF SEQUENCES: 496  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/071,035  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: A. Anders Brookes  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB369P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 366:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 357 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-071-035-366

Query Match 25.0%; Score 53; DB 4; Length 357;  
 Best Local Similarity 36.1%; Pred. No. 6.8;  
 Matches 13; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

QY 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEG 37  
 | :| | ||: |||: : :|| | ||  
 Db 223 MYMANEESAVAVTFSGEAAEMLENNEHLHYVIPSEG 258

# RESULT 32

US-09-252-991A-31873  
 ; Sequence 31873, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31873  
; LENGTH: 452  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31873

Query Match 24.8%; Score 52.5; DB 4; Length 452;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 15; Conservative 3; Mismatches 5; Indels 7; Gaps 2;

Qy 12 AMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|| | |:||| :| | | :| |||  
Db 21 AMLF-GRKSRVE-----SAAKDEDLAWR 43

RESULT 33  
5290690-10  
;Patent No. 5290690  
; APPLICANT: MRABET, NADIR;LASTERS, IGNACE;STANSSENS, PATRICK  
;MATTHYSSENS, GASTON;WODAK, SHOSHANA;QUAX, WILHELMUS J.  
; TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE  
;STABILITY OF PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/398,706  
; FILING DATE: 25-AUG-1989  
;SEQ ID NO:10:  
; LENGTH: 334  
5290690-10

Query Match 24.1%; Score 51; DB 6; Length 334;  
Best Local Similarity 35.0%; Pred. No. 13;  
Matches 14; Conservative 8; Mismatches 10; Indels 8; Gaps 2;

Qy 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEG-----LAW 40  
|| ||: |:| |: : :||| ||| ::  
Db 275 SEEPLVSGDYNGNKN---SSTIDALSTMVMEGSMVKVISW 311

RESULT 34  
5290690-9  
;Patent No. 5290690  
; APPLICANT: MRABET, NADIR;LASTERS, IGNACE;STANSSENS, PATRICK  
;MATTHYSSENS, GASTON;WODAK, SHOSHANA;QUAX, WILHELMUS J.  
; TITLE OF INVENTION: METHODS AND MEANS FOR CONTROLLING THE  
;STABILITY OF PROTEINS  
; NUMBER OF SEQUENCES: 22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/398,706  
; FILING DATE: 25-AUG-1989  
;SEQ ID NO:9:  
; LENGTH: 335  
5290690-9

Query Match 24.1%; Score 51; DB 6; Length 335;

Best Local Similarity 35.0%; Pred. No. 13;  
Matches 14; Conservative 8; Mismatches 10; Indels 8; Gaps 2;

Qy 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEG-----LAW 40  
|| ||: ||: ||: : : || | || : :  
Db 276 SEEPLVSGDYNGNKN---SSTIDALSTMVMEGSMVKVISW 312

RESULT 35

US-09-598-747-27

; Sequence 27, Application US/09598747  
; Patent No. 6531648  
; GENERAL INFORMATION:  
; APPLICANT: Lanahan, Michael B.  
; APPLICANT: Desai, Nalini M.  
; APPLICANT: Gasdaska, Pamela Y.  
; TITLE OF INVENTION: GRAIN PROCESSING METHOD AND TRANSGENIC PLANTS USEFUL  
; TITLE OF INVENTION: THEREIN  
; FILE REFERENCE: A-31383P1  
; CURRENT APPLICATION NUMBER: US/09/598,747  
; CURRENT FILING DATE: 2000-06-21  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-598-747-27

Query Match 23.1%; Score 49; DB 4; Length 310;  
Best Local Similarity 38.7%; Pred. No. 23;  
Matches 12; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

Qy 4 SISENSLVAMDFSGQKSRVIENPTEALSVAV 34  
|| : : ||| : || : | : ||  
Db 88 SIISSETVTAVDFSARPFRVSDSTTVLADAV 118

RESULT 36

US-09-252-991A-17604

; Sequence 17604, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17604  
; LENGTH: 399



; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17604

Query Match 23.1%; Score 49; DB 4; Length 399;  
Best Local Similarity 48.4%; Pred. No. 33;  
Matches 15; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

Qy 11 VAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
||: || | ||| || |||| |  
Db 166 VALLVRGQAERRQRQAGEALQVAFGEGLAAR 196

RESULT 37

US-09-328-352-6943  
; Sequence 6943, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6943  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6943

Query Match 23.1%; Score 49; DB 4; Length 443;  
Best Local Similarity 36.1%; Pred. No. 38;  
Matches 13; Conservative 3; Mismatches 14; Indels 6; Gaps 1;

Qy 10 LVAMDFSGQKSRVIEN-----PTEALSVAVEEGLA 39  
| :|| |: |||| | | | | :|  
Db 304 LYGLDFRGRSKAVIENFTQLNIPLEKLPAYVRHAIA 339

RESULT 38

US-09-328-352-4244  
; Sequence 4244, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4244  
; LENGTH: 133  
; TYPE: PRT

; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4244

Query Match 22.6%; Score 48; DB 4; Length 133;  
Best Local Similarity 30.3%; Pred. No. 10;  
Matches 10; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

Qy 3 RSISENSLVAMDFSGQKSRVIENPTEALSVAVE 35  
| :|::: :| :| :| :|  
Db 68 RPDSDNAVIQIDVYATDEDEVVEQVAESLQFAIE 100

RESULT 39

US-08-454-267-7

; Sequence 7, Application US/08454267

; Patent No. 5843739

; GENERAL INFORMATION:

; APPLICANT: SLABAS, ANTONI R.

; APPLICANT: BROWN, ADRIAN P.

; TITLE OF INVENTION: DNA ENCODING 2-ACYLTRANSFERASES

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600

; CITY: WASHINGTON

; STATE: DC

; COUNTRY: US

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/454,267

; FILING DATE: 08-JUN-1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB93/02528

; FILING DATE: 10-DEC-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: REED, GRANT E.

; REGISTRATION NUMBER: P-41,264

; REFERENCE/DOCKET NUMBER: 0623.0310000/JAG/GER

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 295 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-454-267-7

Query Match 22.6%; Score 48; DB 2; Length 295;  
Best Local Similarity 29.7%; Pred. No. 31;

Matches 11; Conservative 10; Mismatches 12; Indels 4; Gaps 1;

Qy 4 SISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
:: : | :||| : | :||| :||  
Db 200 ALLDKHIAADTFAGQKEQNIGRPIKSLAVV----LSW 232

RESULT 40

US-08-941-319-7

; Sequence 7, Application US/08941319

; Patent No. 5945323

; GENERAL INFORMATION:

; APPLICANT: SLABAS, ANTONI R.

; APPLICANT: BROWN, ADRIAN P.

; TITLE OF INVENTION: DNA ENCODING 2-ACYLTRANSFERASES

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.

; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600

; CITY: WASHINGTON

; STATE: DC

; COUNTRY: US

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/941,319

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/454,267

; FILING DATE: 08-JUN-1995

; APPLICATION NUMBER: PCT/GB93/02528

; FILING DATE: 10-DEC-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: REED, GRANT E.

; REGISTRATION NUMBER: P-41,264

; REFERENCE/DOCKET NUMBER: 0623.0310000/JAG/GER

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 295 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-941-319-7

Query Match 22.6%; Score 48; DB 2; Length 295;

Best Local Similarity 29.7%; Pred. No. 31;

Matches 11; Conservative 10; Mismatches 12; Indels 4; Gaps 1;

Qy 4 SISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40

Db           :: : : | | : | | : | : | : | : |  
200 ALLDKHIAADTFAGQKEQNIGRPIKSLAVV-----LSW 232

Search completed: January 13, 2004, 16:23:27  
Job time : 11.1575 secs

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:19:27 ; Search time 9.48032 Seconds  
(without alignments)  
436.194 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	169	79.7	548	2	I39175	SH2-domain protein
2	162	76.4	532	2	JC5412	epidermal growth f
3	161	75.9	621	2	I49199	growth factor rece
4	159	75.0	535	2	C46243	epidermal growth f
5	66	31.1	655	2	H96692	probable receptor
6	61.5	29.0	404	2	AB2188	hypothetical prote
7	59	27.8	641	2	D87269	hypothetical prote
8	58.5	27.6	685	2	S64158	hypothetical prote
9	56.5	26.7	1346	2	T17412	polyketide synthas
10	55.5	26.2	668	2	S59786	hypothetical prote
11	55	25.9	231	2	AD1785	two components res
12	55	25.9	732	2	H83376	1,4-alpha-glucan b
13	54.5	25.7	196	2	C64891	ferripyochelin-bin

14	54.5	25.7	519	2	AF0345	probable exopolyph
15	54	25.5	231	2	AE1409	two components res
16	54	25.5	301	2	F69045	imidazoleglycerol-
17	54	25.5	432	2	A72477	probable enolase A
18	54	25.5	513	2	G83725	GMP synthetase gua
19	53	25.0	234	2	B95255	glycerol uptake fa
20	53	25.0	579	2	S51528	D-lactate dehydrog
21	52.5	24.8	524	2	AB0158	probable ABC trans
22	52.5	24.8	590	2	F91282	hypothetical prote
23	52.5	24.8	596	2	H86123	hypothetical prote
24	52.5	24.8	967	2	A64710	type III restricti
25	52	24.5	160	2	D82738	hypothetical prote
26	52	24.5	292	2	F83755	hypothetical prote
27	52	24.5	336	2	D64689	quinolinate synthe
28	52	24.5	431	2	G83404	probable chemotaxi
29	52	24.5	502	2	T13620	hypothetical prote
30	52	24.5	519	2	JC4762	RNA-directed RNA p
31	52	24.5	765	2	S76795	hypothetical prote
32	51.5	24.3	185	2	F71349	probable transcrip
33	51.5	24.3	269	2	G75148	hypothetical prote
34	51.5	24.3	324	2	E72536	probable oligopept
35	51.5	24.3	392	2	D72363	carbamoyl-phosphat
36	51.5	24.3	451	2	B90312	hypothetical prote
37	51.5	24.3	476	1	A29881	ubiquinol-cytochro
38	51.5	24.3	801	1	D70309	ribonucleoside-dip
39	51.5	24.3	814	2	F59430	GTPase regulator a
40	51.5	24.3	1423	2	A86289	probable ABC trans
41	51.5	24.3	1847	2	T18308	probable vitelloge
42	51	24.1	233	2	S67937	glycerol uptake fa
43	51	24.1	234	2	A99520	glycerol uptake fa
44	51	24.1	295	2	AC2357	DNA-methyltransfer
45	51	24.1	335	1	DEBSG	glyceraldehyde-3-p

#### ALIGNMENTS

##### RESULT 1

I39175

SH2-domain protein Grb-IR - human

N;Alternate names: insulin receptor cytoplasmic tail-binding protein Grb-IR

C;Species: Homo sapiens (man)

C;Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 05-Nov-1999

C;Accession: I39175

R;Liu, F.; Roth, R.A.

Proc. Natl. Acad. Sci. U.S.A. 92, 10287-10291, 1995

A;Title: Grb-IR: a SH2-domain containing protein that binds to the insulin receptor and inhibits its function.

A;Reference number: I39175; MUID:96036069; PMID:7479769

A;Accession: I39175

A;Status: preliminary; nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-548 <RES>

A;Cross-references: EMBL:U34355; NID:g1079573; PIDN:AAA88819.1; PID:g1079574

A;Note: cloned by a yeast two-hybrid screen with the insulin receptor cytoplasmic domain as the bait

C;Genetics:

A;Gene: GDB:IRBP

A;Cross-references: GDB:697228

C; Superfamily: pleckstrin repeat homology; SH2 homology

F;447-541/Domain: SH2 homology <SH2B>

Query Match 79.7%; Score 169; DB 2; Length 548;  
Best Local Similarity 75.5%; Score 160; DB 2; Length 548;

Best Local Similarity 76.7%; Pred. No. 8e-15;

Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

Db 377 PVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKR 419

## RESULT 2

JC5412

epidermal growth factor receptor-binding protein GRB-7 - human

C; Species: Homo sapiens (man)

C;Date: 10-Jun-1997 #sequence\_revision 18-Jul-1997 #text\_change 21-Jul-2000  
C;Accession: JG5412

C;Accession: JC5412

R; Kishi, T.; Sasaki, H.; Akiyama, N.; Ishizuka, T.; Sakamoto, H.; Aizawa, S.; Sugimura, T.; Terada, M.

Biochem. Biophys. Res. Commun. 232, 5-9, 1997

A;Title: Molecular cloning of human GRB-7 co-amplified with CAB1 and c-ERBB-2 in primary gastric cancer.

A;Reference number: JC5412; MUID:97236270; PMID:9125150

A;Accession: JC5412

A; Molecule type: mRNA

A;Residues: 1-532 &lt;KIS&gt;

A;Cross-references: DDBJ:D43772; NID:g601890; PIDN:BAA07827.1; PID:g601891

C;Comment: This protein contains a pleckstrin domain which mediates protein-protein interaction during signal transduction.

## C; Genetics:

A; Gene: GDB:GRB7

A;Cross-references: GDB:1297554; OMIM:601522

C; Superfamily: pleckstrin repeat homology

F;231-336/Domain: pleckstrin #status predicted <PLE>

F;432-532/Domain: SH2 #status predicted <SH2>

Query Match 76.4%; Score 162; DB 2; Length 532;  
Best Local Similarity 76.4%

Best Local Similarity 74.4%; Pred. No. 6.9e-14;

Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

Db 363

Db 363 PLRSASDNTLVAMDFSGHAGRVIENPREALSVALEEAQAWRKK 405

### RESULT 3

I49199

growth factor receptor binding protein Grb10 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C;Accession: J49188

C;Accession: I49199

R;Ooi, J.; Yajnik, V.; Immanuel, D.; Gordon, M.; Moskow, J.J.; Buchberg, A.M.; Margolis, R

Margolis, B.

Oncogene 10, 1621-1630, 1995

A;Title: The cloning of Grb10 reveals a new family of SH2 domain proteins.  
A;Reference number: I49199; MUID:95249278; PMID:7731717  
A;Accession: I49199  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-621 <RES>  
A;Cross-references: EMBL:U18996; NID:g841209; PIDN:AAB53687.1; PID:g841210  
C;Genetics:  
A;Gene: Grb10  
C;Superfamily: pleckstrin repeat homology; SH2 homology  
C;Keywords: growth factor receptor  
F;520-614/Domain: SH2 homology <SH2B>

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| | | |:| | |  
Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAQSAALEEGHAWR 490

Query Match 75.0%; Score 159; DB 2; Length 535;  
Best Local Similarity 69.8%; Pred. No. 1.8e-13;  
Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY            1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
               |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Db           366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 408



C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C;Accession: H96692  
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
 Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
 Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
 Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
 Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
 Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
 Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
 Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
 S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,  
 M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
 Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
 S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
 Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: H96692  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-655 <STO>  
 A;Cross-references: GB:AE005173; NID:gl1128390; PIDN:AAG31195.1; GSPDB:GN00141  
 C;Genetics:  
 A;Gene: T4024.8  
 A;Map position: 1

Query Match 31.1%; Score 66; DB 2; Length 655;  
 Best Local Similarity 30.2%; Pred. No. 1;  
 Matches 16; Conservative 7; Mismatches 16; Indels 14; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENP-----TEALSVAVEEGLA 39  
 | : || : || | || : || : | :||| |  
 Db 166 PSLKLEGNSFLLNDFGGSCSRVSNPASRTALNTLESTPSTDNLKIALEDGFA 218

# RESULT 6

AB2188

hypothetical protein alr3057 [imported] - *Nostoc* sp. (strain PCC 7120)

C;Species: *Nostoc* sp. PCC 7120

A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AB2188

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
 Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
 Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
 Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing  
*Cyanobacterium Anabaena* sp. strain PCC 7120.

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AB2188

A;Status: preliminary



submitted to the Protein Sequence Database, May 1996  
 A;Reference number: S64153  
 A;Accession: S64158  
 A;Molecule type: DNA  
 A;Residues: 1-685 <VOL>  
 A;Cross-references: EMBL:Z72666; NID:g1322723; PIDN:CAA96856.1; PID:g1322724;  
 GSPDB:GN00007; MIPS:YGL144c  
 A;Experimental source: strain S288C  
 C;Genetics:  
 A;Gene: MIPS:YGL144c  
 A;Cross-references: SGD:S0003112  
 A;Map position: 7L  
 C;Superfamily: conserved hypothetical protein YGL144c

Query Match 27.6%; Score 58.5; DB 2; Length 685;  
 Best Local Similarity 37.8%; Pred. No. 11;  
 Matches 14; Conservative 7; Mismatches 15; Indels 1; Gaps 1;

QY 7 ENSLVAMDFSGQKSRV-IENPTEALSVAVEEGLAWRK 42  
 :| |: |:| | | |:: ||:||||  
 Db 506 KNILLQAFFAGKKERAKYRNLEETIARRWHEGMAWRK 542

#### RESULT 9

T17412

polyketide synthase IV - *Streptomyces venezuelae*

C;Species: *Streptomyces venezuelae*

C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 17-Nov-2000

C;Accession: T17412

R;Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.

Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998

A;Title: A gene cluster for macrolide antibiotic biosynthesis in *streptomyces venezuelae*: architecture of metabolic diversity.

A;Reference number: Z18773; MUID:98445333; PMID:9770448

A;Accession: T17412

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1346 <XUE>

A;Cross-references: EMBL:AF079138; NID:g3808326; PID:g3800837; PIDN:AAC69332.1

C;Genetics:

A;Gene: pikAIV

C;Superfamily: acyl carrier protein homology

C;Keywords: antibiotic biosynthesis; carrier protein

F;945-1016/Domain: acyl carrier protein homology <ACP>

Query Match 26.7%; Score 56.5; DB 2; Length 1346;  
 Best Local Similarity 32.7%; Pred. No. 46;  
 Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
 |:| |:| |:| |:| :| :|:| |:| :|:| |  
 Db 972 PLREIGFDSLTAVDNRNRVRLTGLQLPPTVVFQHPTPVALAERISDELAER 1023

#### RESULT 10

S59786

hypothetical protein YDR320c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein D9798.10  
 C;Species: *Saccharomyces cerevisiae*  
 C;Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 19-Apr-2002  
 C;Accession: S59786  
 R;Du, Z.  
 submitted to the EMBL Data Library, July 1995  
 A;Description: The sequence of *S. cerevisiae* cosmid 9798.  
 A;Reference number: S59418  
 A;Accession: S59786  
 A;Molecule type: DNA  
 A;Residues: 1-668 <DUZ>  
 A;Cross-references: EMBL:U32517; NID:g914989; PID:g914999; GSPDB:GN00004;  
 MIPS:YDR320c  
 A;Experimental source: strain S288C (AB972)  
 C;Genetics:  
 A;Gene: SGD:SWA2; MIPS:YDR320c  
 A;Cross-references: SGD:S0002728  
 A;Map position: 4R

Query Match 26.2%; Score 55.5; DB 2; Length 668;  
 Best Local Similarity 29.5%; Pred. No. 27;  
 Matches 13; Conservative 11; Mismatches 19; Indels 1; Gaps 1;

QY 1 PMRSISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |:| |: :::| |: |: ||| : || : |:|  
 Db 409 PLRIIALSNIIASQLKIGEYSKSIENSSMALELFPSSKAKWKKNK 452

# RESULT 11

AD1785

two components response regulator homolog lin2826 [imported] - *Listeria innocua*  
 (strain Clip11262)

C;Species: *Listeria innocua*

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C;Accession: AD1785

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.;  
 Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve,  
 E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.;  
 Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.;  
 Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.;  
 Jackson, D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.;  
 Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos,  
 B.; Perez-Diaz, J.C.; Remmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes,  
 N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.

A;Title: Comparative genomics of *Listeria* species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AD1785

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-231 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC98052.1; PID:g16415361; GSPDB:GN00178

A;Experimental source: strain Clip11262

C;Genetics:

A;Gene: lin2826

C;Superfamily: ompR protein; response regulator homology

Query Match 25.9%; Score 55; DB 2; Length 231;  
 Best Local Similarity 27.8%; Pred. No. 9.1;  
 Matches 10; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||| : :: | : :: : || | : | | : |  
 Db 192 SENQALRVNMSNIRRKIEQNPAEPAYILTEVGVG YR 227

# RESULT 12

H83376

1,4-alpha-glucan branching enzyme PA2153 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: H83376

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: H83376

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-732 <STO>

A;Cross-references: GB:AE004642; GB:AE004091; NID:g9948163; PIDN:AAG05541.1; GSPDB:GN00131; PASP:PA2153

A;Experimental source: strain PA01

C;Genetics:

A;Gene: glgB; PA2153

C;Superfamily: 1,4-alpha-glucan branching enzyme

Query Match 25.9%; Score 55; DB 2; Length 732;  
 Best Local Similarity 32.5%; Pred. No. 36;  
 Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
 | | : | : || : : : | | | |  
 Db 431 PNRHGGRENLEAIDFLHHLNQVASETPGALVIAEESTAW 470

# RESULT 13

C64891

ferripyochelin-binding protein homolog bl400 - *Escherichia coli* (strain K-12)

C;Species: *Escherichia coli*

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C;Accession: C64891

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.  
 A;Reference number: A64720; MUID:97426617; PMID:9278503  
 A;Accession: C64891  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-196 <BLAT>  
 A;Cross-references: GB:AE000237; GB:U00096; NID:g1787665; PIDN:AAC74482.1;  
 PID:g1787667; UWGP:b1400  
 A;Experimental source: strain K-12, substrain MG1655  
 C;Superfamily: ferripyochelin binding protein

Query Match 25.7%; Score 54.5; DB 2; Length 196;  
 Best Local Similarity 32.6%; Pred. No. 8.8;  
 Matches 15; Conservative 10; Mismatches 14; Indels 7; Gaps 2;

Qy 5 ISENSLV-AMDFSGQKSR-----VIENPTEALSVAVEEGLAWRKK 43  
 | |||:| | | : :: :| :| : |||:| :  
 Db 109 IGENSIVGASAFVKAKAEMPANYLIVGSPAKAIRELSEQELAWKKQ 154

#### RESULT 14

AF0345

probable exopolyphosphatase (EC 3.6.1.11) [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001

C;Accession: AF0345

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.;  
 Prentice, M.B.; Sebahia, M.; James, K.D.; Churcher, C.; Mungall, K.L.; Baker,  
 S.; Basham, D.; Bentley, S.D.; Brooks, K.; Cerdeno-Tarraga, A.M.; Chillingworth,  
 T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Feltwell, T.; Hamlin, N.;  
 Holroyd, S.; Jagels, K.; Leather, S.; Karlyshev, A.V.; Moule, S.; Oyston,  
 P.C.F.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 Whitehead, S.; Barrell, B.G.

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AF0345

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-519 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC93069.1; PID:g15980806; GSPDB:GN00175

C;Genetics:

A;Gene: ppx

C;Superfamily: exopolyphosphatase

C;Keywords: hydrolase

Query Match 25.7%; Score 54.5; DB 2; Length 519;  
 Best Local Similarity 36.2%; Pred. No. 28;  
 Matches 17; Conservative 7; Mismatches 12; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKS-----RVI--ENPTEALSVAVEE 36  
 | ::||| :|| :: | | | | | |  
 Db 473 PHGYLTQNSLVQLDFEREQAYWDDVVGWKLVIIEEEPDEAAKVAPEE 519

#### RESULT 15

AE1409

two components response regulator homolog lmo2678 [imported] - *Listeria monocytogenes* (strain EGD-e)

C;Species: *Listeria monocytogenes*

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C;Accession: AE1409

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, H.; Brandt, P.; Chakraborty, T.; Charbit, A.; Chetouani, F.; Couve, E.; de Daruvar, A.; Dehoux, P.; Domann, E.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Garcia-Del Portillo, F.; Garrido, P.; Gautier, L.; Goebel, W.; Gomez-Lopez, N.; Hain, T.; Hauf, J.; Jackson, D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mata Vicente, J.; Ng, E.; Nordsiek, G.; Novella, S.; de Pablos, B.; Perez-Diaz, J.C.; Rimmel, B.; Rose, M.; Rusniok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, J.; Cossart, P.

A;Title: Comparative genomics of *Listeria* species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AE1409

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-231 <GLA>

A;Cross-references: GB:NC\_003210; PIDN:CAD00891.1; PID:g16412178; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo2678

C;Superfamily: ompR protein; response regulator homology

Query Match 25.5%; Score 54; DB 2; Length 231;

Best Local Similarity 27.8%; Pred. No. 12;

Matches 10; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

Qy 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41

||| : :: | : :: :|| | : | | : :|

Db 192 SENQALRVNMSNIRRKIEKNPAEPAYILTEVGVGGR 227

RESULT 16

F69045

imidazoleglycerol-phosphate synthase (cyclase) hisF MTH1343 [similarity] - *Methanobacterium thermoautotrophicum* (strain Delta H)

C;Species: *Methanobacterium thermoautotrophicum*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-May-2000

C;Accession: F69045

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: F69045

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-301 <MTH>  
A;Cross-references: GB:AE000897; GB:AE000666; NID:g2622439; PIDN:AAB85821.1;  
PID:g2622450  
A;Experimental source: strain Delta H  
C;Genetics:  
A;Gene: MTH1343  
C;Superfamily: cyclase hisF

Query Match 25.5%; Score 54; DB 2; Length 301;  
Best Local Similarity 36.8%; Pred. No. 17;  
Matches 14; Conservative 7; Mismatches 11; Indels 6; Gaps 2;

Qy 6 SENSLVAMDFSGQKSRVIENPTEA---LSVAVEEGLAW 40  
|: :||:| | | ||| |: : |:| |  
Db 153 SQACVVAID--AKRRYIENPRESDERFIIEVDDGYCW 187

#### RESULT 17

A72477

probable enolase APE2458 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C;Accession: A72477

R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.

A;Reference number: A72450; MUID:99310339; PMID:10382966

A;Accession: A72477

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-432 <KAW>

A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81473.1; PID:g5106162

A;Experimental source: strain K1

C;Genetics:

A;Gene: APE2458

C;Superfamily: enolase

Query Match 25.5%; Score 54; DB 2; Length 432;  
Best Local Similarity 43.3%; Pred. No. 26;  
Matches 13; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 10 LVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
|: :| : |||: | | |||:| |  
Db 97 LIELDGTPNKSRLGGNTTTALSIASRAAA 126

#### RESULT 18

G83725

GMP synthetase guaA [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans



C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C;Accession: G83725  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirama, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*.  
 A;Reference number: A83650; MUID:20512582; PMID:11058132  
 A;Accession: G83725  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-513 <STO>  
 A;Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04326.1; GSPDB:GN00137  
 A;Experimental source: strain C-125  
 C;Genetics:  
 A;Gene: guaA  
 C;Superfamily: GMP synthase (glutamine-hydrolyzing); trpG homology

Query Match 25.5%; Score 54; DB 2; Length 513;  
 Best Local Similarity 35.3%; Pred. No. 32;  
 Matches 12; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVE 35  
 | . :|| :| :|| || ::| | | |  
 Db 1 MEQLSEEMIVVLDFFGGQYNQLITRRIRD LGVYSE 34

#### RESULT 19

B95255

glycerol uptake facilitator protein [imported] - *Streptococcus pneumoniae*  
 (strain TIGR4)

C;Species: *Streptococcus pneumoniae*

C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 22-Oct-2001

C;Accession: B95255

R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidelberg, J.; DeBoy, R.T.; Haft, D.H.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.; Kolonay, J.F.; Nelson, W.C.; Peterson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, E.; Khouri, H.; Wolf, A.M.; Utterback, T.R.; Hansen, C.L.; McDonald, L.A.; Feldblyum, T.V.; Angiuoli, S.; Dickinson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, D.A.; Hollingshead, S.K.; Fraser, C.M.

A;Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.

A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: B95255

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-234 <KUR>

A;Cross-references: GB:AE005672; PIDN:AAK76235.1; PID:g14973694; GSPDB:GN00164; TIGR:SP4SP2184

A;Experimental source: strain TIGR4

C;Genetics:

A;Gene: SP2184

C;Superfamily: glycerol facilitator protein

Query Match 25.0%; Score 53; DB 2; Length 234;  
Best Local Similarity 40.0%; Pred. No. 17;  
Matches 12; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 11 VAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
||: ||: | | : ||: || |  
Db 51 VAVFVSGKLSPAYLNPAVTIGVALKGGLPW 80

RESULT 20

S51528

D-lactate dehydrogenase (cytochrome) (EC 1.1.2.4) - yeast (*Kluyveromyces marxianus* var. *lactis*)

C;Species: *Kluyveromyces marxianus* var. *lactis*, *Candida sphaerica*

C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Oct-1999

C;Accession: S51528

R;Lodi, T.; O'Connor, D.; Goffrini, P.; Ferrero, I.

Mol. Gen. Genet. 244, 622-629, 1994

A;Title: Carbon catabolite repression in *Kluyveromyces lactis*: isolation and characterization of the KLDLD gene encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase.

A;Reference number: S51528; MUID:95058916; PMID:7969031

A;Accession: S51528

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-579 <LOD>

A;Cross-references: EMBL:X71628; NID:g602028; PIDN:CAA50635.1; PID:g602029

A;Note: the source is designated as *Kluyveromyces lactis*

C;Keywords: oxidoreductase

Query Match 25.0%; Score 53; DB 2; Length 579;  
Best Local Similarity 32.4%; Pred. No. 51;  
Matches 11; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

Qy 9 SLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRK 42  
| | : | | :: | : | | | : | : | :  
Db 190 SCVVLDISKYLNKIIQLNKEDLDVVVQGGVPWEE 223

RESULT 21

AB0158

probable ABC transport ATP-binding chain YPO1294 [imported] - *Yersinia pestis* (strain CO92)

C;Species: *Yersinia pestis*

C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C;Accession: AB0158

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.;

Prentice, M.B.; Sebaihia, M.; James, K.D.; Churcher, C.; Mungall, K.L.; Baker,

S.; Basham, D.; Bentley, S.D.; Brooks, K.; Cerdeno-Tarraga, A.M.; Chillingworth,

T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Feltwell, T.; Hamlin, N.;

Holroyd, S.; Jagels, K.; Leather, S.; Karlyshev, A.V.; Moule, S.; Oyston,

P.C.F.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

Whitehead, S.; Barrell, B.G.

Nature 413, 523-527, 2001

A;Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360



R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.  
Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: H86123  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-596 <STO>  
A;Cross-references: GB:AE005174; NID:g12519262; PIDN:AAG59452.1; GSPDB:GN00145; UWGP:Z5865  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: yjgL

Query Match 24.8%; Score 52.5; DB 2; Length 596;  
Best Local Similarity 31.0%; Pred. No. 61;  
Matches 13; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

Qy 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
|::| :: : | | : || | :|:: |::|  
Db 552 MQTILKSEVNVSPFIDQQRLNTLNPPENLRIAIEK-FGWKKK 592

#### RESULT 24

A64710

type III restriction enzyme R protein - *Helicobacter pylori* (strain 26695)

C;Species: *Helicobacter pylori*

C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C;Accession: A64710

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.

A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: A64710

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-967 <TOM>

A;Cross-references: GB:AE000650; GB:AE000511; NID:g2314700; PIDN:AAD08561.1; PID:g2314701; TIGR:HP1521

C;Genetics:

A;Start codon: GTG

Query Match 24.8%; Score 52.5; DB 2; Length 967;  
Best Local Similarity 31.8%; Pred. No. 1.1e+02;

Matches 14; Conservative 10; Mismatches 17; Indels 3; Gaps 1;

Qy 3 RSISENSLVAMDFSG---QKSRVIENPTEALSVAVEEGLAWRKK 43  
: |||:|: :|| :|| |:: : ||| : :|  
Db 552 QEISEHSLIKQEFSAELEKSGVVKKGRYGFLLLETLEGLGFGEK 595

RESULT 25

D82738

hypothetical protein XF0981 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C;Accession: D82738

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: D82738

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-160 <SIM>

A;Cross-references: GB:AE003936; GB:AE003849; NID:g9105908; PIDN:AAF83791.1;

GSPDB:GN00128; XFSC:XF0981

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado, J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.

A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0981

Query Match 24.5%; Score 52; DB 2; Length 160;  
Best Local Similarity 45.8%; Pred. No. 15;  
Matches 11; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

Qy 14 DFSGQKSRVIENPTEALSVAVEEG 37  
|| |: | :|| || |: |  
Db 74 DFKGKYDRQLENVTALGRGVQPG 97

RESULT 26

F83755

hypothetical protein BH0846 [imported] - *Bacillus halodurans* (strain C-125)

C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C;Accession: F83755

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirama, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*.

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: F83755

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-292 <STO>

A;Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04565.1;

GSPDB:GN00137

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH0846

C;Superfamily: hypothetical protein ywpJ

Query Match 24.5%; Score 52; DB 2; Length 292;  
Best Local Similarity 39.4%; Pred. No. 31;  
Matches 13; Conservative 6; Mismatches 10; Indels 4; Gaps 1;

Qy 10 LVAMDFSG---QKSRVIENPTEALSVAVEEGL 38  
|:|:| | :|| : | ||:| | | :  
Db 4 LIAIDLDTLLNEKSTISEENTESLQRAQEAGM 36

RESULT 27

D64689

quinolinate synthetase A - *Helicobacter pylori* (strain 26695)

C;Species: *Helicobacter pylori*

C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C;Accession: D64689

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.

A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: D64689

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-336 <TOM>

A;Cross-references: GB:AE000636; GB:AE000511; NID:g2314517; PIDN:AAD08398.1;

PID:g2314524; TIGR:HP1356

C;Superfamily: *Helicobacter pylori* quinolinate synthetase A

Query Match 24.5%; Score 52; DB 2; Length 336;

Best Local Similarity 34.9%; Pred. No. 36;

Matches 15; Conservative 9; Mismatches 13; Indels 6; Gaps 2;

Qy 7 ENSLVA-MDFSGQKSRVIE-----NPTEALSVAVEEGLAWRKK 43

Db 228 EPSVVSNAADFSGSTSQIIEFVEKLSPNQKVAIGTESHLVNRLK 270

RESULT 28

G83404

probable chemotaxis transducer PA1930 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: G83404

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: G83404

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-431 <STO>

A;Cross-references: GB:AE004619; GB:AE004091; NID:g9947920; PIDN:AAG05318.1;

GSPDB:GN00131; PASP:PA1930

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA1930

Query Match 24.5%; Score 52; DB 2; Length 431;

Best Local Similarity 48.0%; Pred. No. 49;

Matches 12; Conservative 3; Mismatches 4; Indels 6; Gaps 1;

Qy 17 GQKSRVIENPTEALSVAVEEGLAWR 41

Db 4 GRKSRAVE-----SAAKDEDLAWR 22

RESULT 29

T13620

hypothetical protein gp502 - Streptococcus phage phi-Sfill

C;Species: Streptococcus phage phi-Sfill

C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999

C;Accession: T13620

R;Lucchini, S.; Desiere, F.; Bruessow, H.

Virology 246, 63-73, 1998

A;Title: The structural gene module in Streptococcus thermophilus bacteriophage phi Sfill shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts.

A;Reference number: Z17696; MUID:98321150; PMID:9656994

A;Accession: T13620

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-502 <LUC>

A;Cross-references: EMBL:AF057033; NID:g3320432; PID:g3320433; PIDN:AAC34397.1

A;Experimental source: specific\_host Streptococcus thermophilus

Query Match 24.5%; Score 52; DB 2; Length 502;  
Best Local Similarity 33.3%; Pred. No. 58;  
Matches 14; Conservative 6; Mismatches 14; Indels 8; Gaps 1;

Qy 3 RSISENSLVAMDFSGQKSR-----VIENPTEALSVAVEE 36  
:|:| :| ||| :::||| | ||  
Db 424 KSLYEQVSILNDLGGQVSQETALSLSGLVENPTEELDKINEE 465

RESULT 30

JC4762

RNA-directed RNA polymerase (EC 2.7.7.48) - Mycovirus Fusov

C;Species: Mycovirus Fusov

A;Note: host Fusarium solani f.sp. robiniae

C;Date: 10-May-1996 #sequence\_revision 16-Aug-1996 #text\_change 05-Nov-1999

C;Accession: JC4762

R;Nogawa, M.; Kageyama, T.; Nakatani, A.; Taguchi, G.; Shimosaka, M.; Okazaki, M.

Biosci. Biotechnol. Biochem. 60, 784-788, 1996

A;Title: Cloning and characterization of mycovirus double-stranded RNA from the plant pathogenic fungus, Fusarium solani f.sp. robiniae.

A;Reference number: JC4762; MUID:96261063; PMID:8704307

A;Accession: JC4762

A;Molecule type: mRNA

A;Residues: 1-519 <NOG>

A;Cross-references: DDBJ:D55668; NID:g893387; PIDN:BAA09520.1; PID:g893388

A;Note: RNA polymerase

C;Comment: This enzyme is responsible for replication of two segmented double-stranded RNA genomes, M1 and M2.

C;Keywords: nucleotidyltransferase

F;260-269/Region: RNA-directed RNA polymerase motif 1

F;332-360/Region: RNA-directed RNA polymerase motif 2

F;366-375/Region: RNA-directed RNA polymerase motif 3

Query Match 24.5%; Score 52; DB 2; Length 519;  
Best Local Similarity 41.2%; Pred. No. 61;  
Matches 14; Conservative 6; Mismatches 12; Indels 2; Gaps 1;



```

Qy      11 VAMDFSGQKSRVIENPTE--ALSVAVEEGLAWRK 42
      ||| | :|| :|: || | | :| |:|:
Db     393 VGMDLSDEKSI SVEDATELKL LGVRYRDGHAFRE 426

```

RESULT 31

S76795

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C;Species: Synechocystis sp.

A; Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence revision 25-Apr-1997 #text change 08-Oct-1999

C;Accession: S76795

R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. . .

A;Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76795

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-765 <KAN>

A;Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAA18707.1;

PID:d1019440; PID:g1653796

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 24.5%; Score 52; DB 2; Length 765;

Best Local Similarity 26.5%; Pred. No. 96;

Matches 9; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Qy 4 SISENSLVAMDFSGQKSRVIENPTEALSVAVEEG 37  
|:: ::| | | :::: ::||: :||  
Db 83 SLTSSLTLTEDLRGQSTQLVOLTSOALTEPTKEG 116

RESULT 32

F71349

probable transcription antitermination protein (nusG) - syphilis spirochete

C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 24-Jul-1998 #sequence revision 24-Jul-1998 #text change 05-Nov-1999

C;Accession: F71349

R.; Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald, L.; Artiach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A;Title: Complete genome sequence of *Treponema pallidum*, the syphilis spirochete.

A;Reference number: A71250; MUID:98332770; PMID:9665876

A;Accession: F71349

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-185 <COL>  
A;Cross-references: GB:AE001205; GB:AE000520; NID:g3322501; PIDN:AAC65224.1;  
PID:g3322506  
A;Experimental source: strain Nichols  
C;Genetics:  
A;Gene: TP0236  
C;Superfamily: transcription antitermination factor nusG

Query Match 24.3%; Score 51.5; DB 2; Length 185;  
Best Local Similarity 35.9%; Pred. No. 21;  
Matches 14; Conservative 7; Mismatches 13; Indels 5; Gaps 1;

Qy 5 ISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
|:: || |::|::| | ||| ::||  
Db 128 IAQTFVLV-----GQQVRIVEGPFATFSGEVEEVMSE RNK 161

RESULT 33

G75148

hypothetical protein PAB0223 - *Pyrococcus abyssi* (strain Orsay)

C;Species: *Pyrococcus abyssi*

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C;Accession: G75148

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution.

A;Reference number: A75001

A;Accession: G75148

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-269 <KAW>

A;Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49270.1;  
PID:e1515165; PID:g5457780

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB0223

Query Match 24.3%; Score 51.5; DB 2; Length 269;  
Best Local Similarity 41.9%; Pred. No. 33;  
Matches 13; Conservative 7; Mismatches 6; Indels 5; Gaps 1;

Qy 11 VAMDFSGQK-----SRVIENPTEALSVAVEE 36  
|:| :| ||::|: |:|:| ||||  
Db 48 VTIDLPREKKGIHMSRLVESITDAMSEAVEE 78

RESULT 34

E72536

probable oligopeptide transport ATP-binding protein APE1578 - *Aeropyrum pernix* (strain K1)

C;Species: *Aeropyrum pernix*

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C;Accession: E72536

R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.  
 DNA Res. 6, 83-101, 1999  
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum pernix* K1.  
 A;Reference number: A72450; MUID:99310339; PMID:10382966  
 A;Accession: E72536  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-324 <KAW>  
 A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80578.1; PID:g5105265  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: APE1578  
 C;Superfamily: inner membrane protein malk; ATP-binding cassette homology  
 F;25-231/Domain: ATP-binding cassette homology <ABC>

Query Match 24.3%; Score 51.5; DB 2; Length 324;  
 Best Local Similarity 31.9%; Pred. No. 41;  
 Matches 15; Conservative 4; Mismatches 7; Indels 21; Gaps 2;

Qy 17 GQKSRVI-----ENPTEALSVAVE-----EGLAWRK 42  
 ||| ||: : || || ||: : ||| |  
 Db 158 GQKQRVVIAMALALEPDIIVIADEPTTALDVVVQAQILNLLKKLAWEK 204

# RESULT 35

D72363

carbamoyl-phosphate synthetase, small subunit - *Thermotoga maritima* (strain MSB8)

C;Species: *Thermotoga maritima*

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C;Accession: D72363

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*.

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: D72363

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-392 <ARN>

A;Cross-references: GB:AE001730; GB:AE000512; NID:g4981062; PIDN:AAD35643.1; PID:g4981073; TIGR:TM0558

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM0558

C;Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain;  
carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homology; trpG  
homology

F;177-385/Domain: trpG homology <TRG>

Query Match 24.3%; Score 51.5; DB 2; Length 392;  
Best Local Similarity 31.8%; Pred. No. 51;  
Matches 14; Conservative 10; Mismatches 15; Indels 5; Gaps 3;

Qy 2 MRSISEN-SLVAMDFSG---QKSRVIENPTEALSVAV-EEGLAW 40  
:: : |: |:| |:| |::||| || |: |:|  
Db 143 VKRVKESPSIVGRDLAGLVSPKEVIVENPEGDFSVVVLDSGVKW 186

#### RESULT 36

B90312

hypothetical protein SS01531 [imported] - Sulfolobus solfataricus

C;Species: Sulfolobus solfataricus

C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C;Accession: B90312

R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-Weiher, C.C.Y.; Clausen, I.G.; Curtis, B.A.; De Moors, A.; Erauso, G.; Fletcher, C.; Gordon, P.M.K.; Heikamp-de Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.; Schenk, M.E.; Theriault, C.; Tolstrup, N.; Charlebois, R.L.; Doolittle, W.F.; Dugué, M.; Gaasterland, T.; Garrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A;Description: Sulfolobus solfataricus complete genome.

A;Reference number: A99139

A;Accession: B90312

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-451 <KUR>

A;Cross-references: GB:AE006641; NID:g13814763; PIDN:AAK41753.1; GSPDB:GN00155

C;Genetics:

A;Gene: SS01531

Query Match 24.3%; Score 51.5; DB 2; Length 451;  
Best Local Similarity 45.2%; Pred. No. 60;  
Matches 14; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

Qy 8 NSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38  
|: |:| |:| |:| |::||| || |: |:|  
Db 86 NNCVILDLS-KLNRIIEFNEDLSVTVEVGI 115

#### RESULT 37

A29881

ubiquinol-cytochrome-c reductase (EC 1.10.2.2) beta chain precursor - Neurospora crassa

N;Alternate names: beta-MPP; mitochondrial processing peptidase enhancing protein; PEP; ubiquinol-cytochrome-c reductase (EC 1.10.2.2) core protein I

C;Species: Neurospora crassa

C;Date: 31-Dec-1993 #sequence\_revision 14-Jul-1994 #text\_change 03-Jun-2002

C;Accession: A29881; B29881; S03968

R;Hawltitschek, G.; Schneider, H.; Schmidt, B.; Tropschug, M.; Hartl, F.U.; Neupert, W.

Cell 53, 795-806, 1988

A;Title: Mitochondrial protein import: identification of processing peptidase and of PEP, a processing enhancing protein.

A;Reference number: A29881; MUID:88223372; PMID:2967109

A;Accession: A29881

A;Molecule type: mRNA

A;Residues: 1-476 <HAW>

A;Cross-references: EMBL:M20928; NID:g168857; PIDN:AAA33606.1; PID:g168858

A;Accession: B29881

A;Molecule type: protein

A;Residues: 'XX',31-34 <HA2>

R;Schulte, U.; Arretz, M.; Schneider, H.; Tropschug, M.; Wachter, E.; Neupert, W.; Weiss, H.

Nature 339, 147-149, 1989

A;Title: A family of mitochondrial proteins involved in bioenergetics and biogenesis.

A;Reference number: S03968; MUID:89238559; PMID:2524007

A;Accession: S03968

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-476 <SCH>

A;Cross-references: EMBL:M20928; NID:g168857; PIDN:AAA33606.1; PID:g168858

A;Note: part of this sequence was confirmed by protein sequencing

C;Comment: In *Neurospora crassa* the beta chain of the mitochondrial processing peptidase and the core I protein of ubiquinol-cytochrome-c reductase are identical. The protein is bifunctional and participates both in protein processing and electron transport.

C;Superfamily: mitochondrial processing peptidase alpha chain

C;Keywords: heterodimer; hydrolase; metalloproteinase; mitochondrial matrix; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory chain

F;1-28/Domain: transit peptide (mitochondrion) #status predicted <TNP>

F;29-476/Product: mitochondrial processing peptidase beta chain #status experimental <MAT>

Query Match 24.3%; Score 51.5; DB 1; Length 476;  
Best Local Similarity 35.7%; Pred. No. 64;  
Matches 15; Conservative 9; Mismatches 15; Indels 3; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSRVIEN--PTEALSVAVEEGLAW 40  
|: || | || | |: :: || ::|| ||::|  
Db 251 PVSSASILSKKKPDFIGSDIRDDTIPTANIAIAV-EGVSW 291

RESULT 38

D70309

ribonucleoside-diphosphate reductase (EC 1.17.4.1) alpha chain [similarity] -  
*Aquifex aeolicus*

C;Species: *Aquifex aeolicus*

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C;Accession: D70309

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.

Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: D70309  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-801 <AQF>  
 A;Cross-references: GB:AE000673; NID:g2982834; PIDN:AAC06460.1; PID:g2982838;  
 GB:AE000657  
 A;Experimental source: strain VF5  
 C;Genetics:  
 A;Gene: nrdA  
 C;Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain  
 C;Keywords: deoxyribonucleotide biosynthesis; oxidoreductase; redox-active  
 disulfide  
 F;235-521,796-799/Disulfide bonds: redox-active #status predicted  
 F;483,487/Active site: Asn, Glu #status predicted  
 F;485/Active site: Cys (cysteine thiyl radical intermediate) #status predicted

Query Match 24.3%; Score 51.5; DB 1; Length 801;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
 Matches 15; Conservative 5; Mismatches 6; Indels 9; Gaps 2;

QY 18 QKSRVIENPTE-----ALSVAV-----EEGLAWRKK 43  
 :: |||| | | |::|| | | | |  
 Db 171 EEGRVIELPQEMYMLIAMTLAVPEKPEERLKWAKK 205

# RESULT 39

F59430

GTPase regulator associated with focal adhesion kinase pp125 [imported] - human

C;Species: Homo sapiens (man)

C;Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 23-Sep-2002

C;Accession: F59430; G59430; H59430

R;Taylor, J.M.; Macklem, M.M.; Parsons, J.T.

J. Cell. Sci. 112 (Pt 2), 231-242, 1999

A;Title: Cytoskeletal changes induced by GRAF, the GTPase regulator associated with focal adhesion kinase, are mediated by Rho.

A;Reference number: F59430

A;Accession: F59430

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-814 <TAY>

A;Cross-references: GB:NP\_055886; PID:g7662208; PIDN:NP\_055886.1

R;Borkhardt, A.; Bojesen, S.; Haas, O.A.; Fuchs, U.; Bartelheimer, D.;

Loncarevic, I.F.; Bohle, R.M.; Harbott, J.; Repp, R.; Jaeger, U.; Viehmann, S.;

Henn, T.; Korth, P.; Scharr, D.; Lampert, F.

Proc. Natl. Acad. Sci. U.S.A. 97, 9168-9173, 2000

A;Title: The human GRAF gene is fused to MLL in a unique t(5;11)(q31;q23) and both alleles are disrupted in three cases of myelodysplastic syndrome/acute myeloid leukemia with a deletion 5q.

A;Reference number: G59430

A;Accession: G59430

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-814 <BOR>

A;Cross-references: GB:NP\_055886; PID:g7662208; PIDN:NP\_055886.1

R;Xia, J.H.; Tang, X.X.; Yu, K.P.; Pan, Q.; Dai, H.P.

submitted to GenBank, April 2002



Qy            7 ENSLVAMDFSGQK-SRVIENTEALSV 32  
              :|::| :| |    || :|| :| | |  
Db            710 QNAILANEFFGHSWSRAVENSSETLGV 736

Search completed: January 13, 2004, 16:24:10  
Job time : 12.4803 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:22:54 ; Search time 18.622 Seconds  
(without alignments)  
465.304 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query				
No.	Score	Match Length	DB	ID		Description

1	212	100.0	540	15	US-10-242-332-2	Sequence 2, Appli
2	212	100.0	540	16	US-10-323-001-2	Sequence 2, Appli
3	162	76.4	532	15	US-10-097-340-125	Sequence 125, App
4	162	76.4	532	15	US-10-233-098-2	Sequence 2, Appli
5	161	75.9	621	15	US-10-242-332-4	Sequence 4, Appli
6	161	75.9	621	16	US-10-323-001-4	Sequence 4, Appli
7	159	75.0	375	12	US-10-094-749-3245	Sequence 3245, Ap
8	159	75.0	535	15	US-10-242-332-3	Sequence 3, Appli
9	159	75.0	535	16	US-10-323-001-3	Sequence 3, Appli
10	59.5	28.1	1346	11	US-09-793-708-4	Sequence 4, Appli
11	59.5	28.1	1346	12	US-10-201-365-5	Sequence 5, Appli
12	59.5	28.1	1346	12	US-10-160-539-4	Sequence 4, Appli
13	58	27.4	1024	15	US-10-211-962-85	Sequence 85, Appl
14	57.5	27.1	537	14	US-10-037-667-1	Sequence 1, Appli
15	56.5	26.7	1346	10	US-09-861-289-37	Sequence 37, Appl
16	56.5	26.7	1346	10	US-09-860-846-37	Sequence 37, Appl
17	56.5	26.7	1346	11	US-09-988-384B-37	Sequence 37, Appl
18	56.5	26.7	1346	11	US-09-836-821-37	Sequence 37, Appl
19	56.5	26.7	1346	12	US-10-271-889-37	Sequence 37, Appl
20	56.5	26.7	11877	10	US-09-861-289-6	Sequence 6, Appli
21	56.5	26.7	11877	10	US-09-860-846-6	Sequence 6, Appli
22	56.5	26.7	11877	11	US-09-836-821-6	Sequence 6, Appli
23	56.5	26.7	11877	12	US-10-271-889-49	Sequence 49, Appl
24	56.5	26.7	12199	11	US-09-988-384B-6	Sequence 6, Appli
25	54.5	25.7	196	12	US-10-287-274-379	Sequence 379, App
26	54	25.5	301	12	US-10-369-493-1220	Sequence 1220, Ap
27	54	25.5	432	12	US-10-369-493-22965	Sequence 22965, A
28	53	25.0	234	11	US-09-769-787-9	Sequence 9, Appli
29	53	25.0	530	10	US-09-738-626-4780	Sequence 4780, Ap
30	52	24.5	208	12	US-10-326-671-236	Sequence 236, App
31	52	24.5	330	9	US-09-815-242-10691	Sequence 10691, A
32	51.5	24.3	392	12	US-10-369-493-2956	Sequence 2956, Ap
33	51.5	24.3	617	12	US-10-369-493-3798	Sequence 3798, Ap
34	51.5	24.3	801	12	US-10-369-493-11	Sequence 11, Appl
35	51	24.1	335	12	US-10-369-493-23303	Sequence 23303, A
36	51	24.1	431	12	US-10-032-585-7696	Sequence 7696, Ap
37	50.5	23.8	406	12	US-10-369-493-11560	Sequence 11560, A
38	50.5	23.8	450	12	US-10-369-493-14587	Sequence 14587, A
39	50.5	23.8	465	12	US-10-369-493-14939	Sequence 14939, A
40	50.5	23.8	487	12	US-10-369-493-14160	Sequence 14160, A
41	50.5	23.8	638	14	US-10-072-621-10	Sequence 10, Appl
42	50.5	23.8	674	15	US-10-090-455-4	Sequence 4, Appli
43	50	23.6	120	10	US-09-738-626-5486	Sequence 5486, Ap
44	50	23.6	366	12	US-10-369-493-8582	Sequence 8582, Ap
45	49.5	23.3	257	10	US-09-902-525-35	Sequence 35, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-242-332-2

; Sequence 2, Application US/10242332

; Publication No. US20030044834A1

; GENERAL INFORMATION:

; APPLICANT: Daly, Roger John



Db 367 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 409

RESULT 3

US-10-097-340-125

; Sequence 125, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN

; APPLICANT: Manjula GANNAVAPU

; APPLICANT: Sebastian HOERSCH

; APPLICANT: Shubhangi KAMATKAR

; APPLICANT: Steve G. KOVATS

; APPLICANT: Rachel E. MEYERS

; APPLICANT: Michael MORRISEY

; APPLICANT: Peter OLANDT

; APPLICANT: Ami SEN

; APPLICANT: Peter VEIBY

; APPLICANT: Gordon B. MILLS

; APPLICANT: Robert C. BAST, Jr.

; APPLICANT: Karen LU

; APPLICANT: Rosemarie SCHMANDT

; APPLICANT: Xumei ZHAO

; APPLICANT: Karen GLATT

; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,

; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer

; FILE REFERENCE: MRI-030

; CURRENT APPLICATION NUMBER: US/10/097,340

; CURRENT FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: 60/276,025

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/325,149

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/276,026

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/324,967

; PRIOR FILING DATE: 2001/09/26

; PRIOR APPLICATION NUMBER: 60/311,732

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: 60/325,102

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/323,580

; PRIOR FILING DATE: 2001-09-19

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 125

; LENGTH: 532

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-097-340-125

Query Match 76.4%; Score 162; DB 15; Length 532;

Best Local Similarity 74.4%; Pred. No. 1.4e-14;

Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

Db 363 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEEAQAWRKK 405

RESULT 4

US-10-233-098-2

; Sequence 2, Application US/10233098

; Publication No. US20030109440A1

; GENERAL INFORMATION:

; APPLICANT: Chu, Peter

; APPLICANT: Li, Congfen

; APPLICANT: Liao, X. Charlene

; APPLICANT: Masuda, Esteban

; APPLICANT: Pardo, Jorge

; APPLICANT: Zhao, Haoran

; APPLICANT: Rigel Pharmaceuticals, Incorporated

; TITLE OF INVENTION: GRB7: No. US20030109440A1el Regulator of Lymphocytic Signaling

; FILE REFERENCE: 021044-004500

; CURRENT APPLICATION NUMBER: US/10/233,098

; CURRENT FILING DATE: 2002-08-30

; PRIOR APPLICATION NUMBER: US 60/327,212

; PRIOR FILING DATE: 2001-10-03

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 532

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human wild-type growth factor receptor-bound 7

; OTHER INFORMATION: (GRB7)

US-10-233-098-2

Query Match 76.4%; Score 162; DB 15; Length 532;

Best Local Similarity 74.4%; Pred. No. 1.4e-14;

Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43

Db 363 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEEAQAWRKK 405

RESULT 5

US-10-242-332-4

; Sequence 4, Application US/10242332

; Publication No. US20030044834A1

; GENERAL INFORMATION:

; APPLICANT: Daly, Roger John

; APPLICANT: Sutherland, Robert Lyndsay

; TITLE OF INVENTION: GDU, A novel signalling protein

; FILE REFERENCE: 273402001710

; CURRENT APPLICATION NUMBER: US/10/242,332

; CURRENT FILING DATE: 2002-09-11

; PRIOR APPLICATION NUMBER: US 08/945,771

; PRIOR FILING DATE: 1998-04-22

; PRIOR APPLICATION NUMBER: PCT/AU96/00258

; PRIOR FILING DATE: 1996-05-02  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 621  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-242-332-4

Query Match 75.9%; Score 161; DB 15; Length 621;  
Best Local Similarity 78.0%; Pred. No. 2.4e-14;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| | | |:| | |  
Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAQSAALEEGHAWR 490

RESULT 6

US-10-323-001-4  
; Sequence 4, Application US/10323001  
; Publication No. US20030129639A1  
; GENERAL INFORMATION:  
; APPLICANT: Daly, Roger John  
; APPLICANT: Sutherland, Robert Lyndsay  
; TITLE OF INVENTION: GDU, A novel signalling protein  
; FILE REFERENCE: 273402001710  
; CURRENT APPLICATION NUMBER: US/10/323,001  
; CURRENT FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: US/10/242,332  
; PRIOR FILING DATE: 2002-09-11  
; PRIOR APPLICATION NUMBER: US 08/945,771  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: PCT/AU96/00258  
; PRIOR FILING DATE: 1996-05-02  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 621  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-323-001-4

Query Match 75.9%; Score 161; DB 16; Length 621;  
Best Local Similarity 78.0%; Pred. No. 2.4e-14;  
Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
|||:||||||| |||:| | | |:| | |  
Db 450 PMRSVSENSLVAMDFSGQIGRVIDNPAAQSAALEEGHAWR 490

RESULT 7

US-10-094-749-3245  
; Sequence 3245, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:



```
; PRIOR FILING DATE: 1996-05-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-242-332-3
```

Query Match 75.0%; Score 159; DB 15; Length 535;  
Best Local Similarity 69.8%; Pred. No. 3.9e-14;  
Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy            1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
             |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
Db           366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 408

## RESULT 9

US-10-323-001-3

```

; Sequence 3, Application US/10323001
; Publication No. US20030129639A1
; GENERAL INFORMATION:
; APPLICANT: Daly, Roger John
; APPLICANT: Sutherland, Robert Lyndsay
; TITLE OF INVENTION: GDU, A novel signalling protein
; FILE REFERENCE: 273402001710
; CURRENT APPLICATION NUMBER: US/10/323,001
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US/10/242,332
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 08/945,771
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: PCT/AU96/00258
; PRIOR FILING DATE: 1996-05-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-323-001-3

```

Query Match 75.0%; Score 159; DB 16; Length 535;  
Best Local Similarity 69.8%; Pred. No. 3.9e-14;  
Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY           1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
             | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db          366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK 408

RESULT 10

US-09-793-708-4

```
; Sequence 4, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
```



; APPLICANT: ASHLEY, Gary  
 ; APPLICANT: BETLACH, Melanie C.  
 ; APPLICANT: BETLACH, Mary C.  
 ; APPLICANT: McDANIEL, Robert  
 ; APPLICANT: TANG, Li  
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
 ; FILE REFERENCE: 300622002121  
 ; CURRENT APPLICATION NUMBER: US/09/793,708  
 ; CURRENT FILING DATE: 2001-02-22  
 ; PRIOR APPLICATION NUMBER: US 09/657,440  
 ; PRIOR FILING DATE: 2000-09-07  
 ; PRIOR APPLICATION NUMBER: US 09/320,878  
 ; PRIOR FILING DATE: 1999-05-27  
 ; PRIOR APPLICATION NUMBER: US 09/141,908  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: US 09/073,538  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/846,247  
 ; PRIOR FILING DATE: 1997-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/134,990  
 ; PRIOR FILING DATE: 1999-05-20  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1346  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-793-708-4

Query Match 28.1%; Score 59.5; DB 11; Length 1346;  
 Best Local Similarity 34.6%; Pred. No. 23;  
 Matches 18; Conservative 9; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
 |:| :| |:| :| | :| | :| :| :| :|  
 Db 972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023

# RESULT 11

US-10-201-365-5

; Sequence 5, Application US/10201365  
 ; Publication No. US20030148469A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ASHLEY, Gary  
 ; APPLICANT: BETLACH, Melanie C.  
 ; APPLICANT: BETLACH, Mary  
 ; APPLICANT: MCDANIEL, Robert  
 ; APPLICANT: TANG, Li  
 ; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A  
 MODULAR  
 ; TITLE OF INVENTION: PKS GENE CLUSTER AS SCAFFOLD  
 ; FILE REFERENCE: 300622002103  
 ; CURRENT APPLICATION NUMBER: US/10/201,365  
 ; CURRENT FILING DATE: 2002-07-22  
 ; PRIOR APPLICATION NUMBER: US 09/141,908  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: US 09/073,538

; PRIOR FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-10-201-365-5

Query Match 28.1%; Score 59.5; DB 12; Length 1346;  
Best Local Similarity 34.6%; Pred. No. 23;  
Matches 18; Conservative 9; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : | | |:| | |: : : || |  
Db 972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023

RESULT 12

US-10-160-539-4

; Sequence 4, Application US/10160539  
; Publication No. US20030162262A1  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: McDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/10/160,539  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: US/09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-10-160-539-4

Query Match 28.1%; Score 59.5; DB 12; Length 1346;  
Best Local Similarity 34.6%; Pred. No. 23;  
Matches 18; Conservative 9; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : | | |:| | |: : : || |  
Db 972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFEHPTPVALAERISDELAER 1023

RESULT 13

US-10-211-962-85

; Sequence 85, Application US/10211962  
 ; Publication No. US20030082640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Herz, Joachim  
 ; APPLICANT: Gotthardt, Michael  
 ; TITLE OF INVENTION: LDL Receptor Signaling Pathways  
 ; FILE REFERENCE: UTSW0708  
 ; CURRENT APPLICATION NUMBER: US/10/211,962  
 ; CURRENT FILING DATE: 2002-08-01  
 ; PRIOR APPLICATION NUMBER: US/09/562,737  
 ; PRIOR FILING DATE: 2000-05-01  
 ; NUMBER OF SEQ ID NOS: 132  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 85  
 ; LENGTH: 1024  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 ; OTHER INFORMATION: Sequence  
 US-10-211-962-85

Query Match 27.4%; Score 58; DB 15; Length 1024;  
 Best Local Similarity 27.9%; Pred. No. 27;  
 Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 1;

QY 3 RSISENSLVAMDFSGQ--KSRVIENPTEALSVAVEEGLAWRKK 43  
 : : : || : |||| : : : || : : : || :  
 Db 460 QAVAANSAASRDFSGQGGLGELLESRSEASKLSSKTAKWRNR 502

#### RESULT 14

US-10-037-667-1

; Sequence 1, Application US/10037667  
 ; Publication No. US20020177145A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morgan, Bruce A.  
 ; TITLE OF INVENTION: REGULATION OF NEURAL DEVELOPMENT BY  
 ; TITLE OF INVENTION: DAEDALOS  
 ; FILE REFERENCE: 10287-044001  
 ; CURRENT APPLICATION NUMBER: US/10/037,667  
 ; CURRENT FILING DATE: 2002-07-23  
 ; PRIOR APPLICATION NUMBER: 60/243,110  
 ; PRIOR FILING DATE: 2000-10-25  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 537  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus

US-10-037-667-1

Query Match 27.1%; Score 57.5; DB 14; Length 537;  
 Best Local Similarity 36.4%; Pred. No. 14;  
 Matches 16; Conservative 8; Mismatches 15; Indels 5; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEAL----SVAVEEGLA 39

Db                   | ||:| ||:   :| ::| : | ||       || ||: |:       37 PSRSLSANSIKVEMYSDEESSRLLGPDERLLDKDDSVIVEDSLS 80

RESULT 15

US-09-861-289-37

; Sequence 37, Application US/09861289  
 ; Patent No. US20020110897A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/861,289  
 ; CURRENT FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 37  
 ; LENGTH: 1346  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-861-289-37

Query Match                   26.7%; Score 56.5; DB 10; Length 1346;  
 Best Local Similarity       32.7%; Pred. No. 61;  
 Matches   17; Conservative   10; Mismatches   14; Indels   11; Gaps   2;

Qy                   1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
                   |:| | :|| |:| | : :|                   | :|| | | : : || |  
 Db                   972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVFQHPPTPVALAERISDELAER 1023

RESULT 16

US-09-860-846-37

; Sequence 37, Application US/09860846  
 ; Patent No. US20020164742A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/860,846  
 ; CURRENT FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 37  
 ; LENGTH: 1346  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces venezuelae

US-09-860-846-37

Query Match 26.7%; Score 56.5; DB 10; Length 1346;  
Best Local Similarity 32.7%; Pred. No. 61;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | :| | :|| | |:| | |:| : : || |  
Db 972 PLREIGFDSLTAVD FRNRVNRLTGLQLPPTVVFQHPTPVALAERISDELAER 1023

RESULT 17

US-09-988-384B-37

; Sequence 37, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 37  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae

US-09-988-384B-37

Query Match 26.7%; Score 56.5; DB 11; Length 1346;  
Best Local Similarity 32.7%; Pred. No. 61;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | :| | :|| | |:| | |:| : : || |  
Db 972 PLREIGFDSLTAVD FRNRVNRLTGLQLPPTVVFQHPTPVALAERISDELAER 1023

RESULT 18

US-09-836-821-37

; Sequence 37, Application US/09836821  
; Publication No. US20030087405A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/836,821  
; CURRENT FILING DATE: 2001-04-17

; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 37  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-836-821-37

Query Match 26.7%; Score 56.5; DB 11; Length 1346;  
Best Local Similarity 32.7%; Pred. No. 61;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : | | :|| | |: : : || |  
Db 972 PLREIGFDSLTAVDNRNRVRLTGLQLPPTVVFQHPTPVALAERISDELAER 1023

RESULT 19

US-10-271-889-37

; Sequence 37, Application US/10271889  
; Publication No. US20030194784A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin  
; FILE REFERENCE: 600.582US1  
; CURRENT APPLICATION NUMBER: US/10/271,889  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 09/861,289  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 09/860,846  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 09/836,821  
; PRIOR FILING DATE: 2001-04-17  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 37  
; LENGTH: 1346  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-10-271-889-37

Query Match 26.7%; Score 56.5; DB 12; Length 1346;  
Best Local Similarity 32.7%; Pred. No. 61;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | : | | :|| | |: : : || |  
Db 972 PLREIGFDSLTAVDNRNRVRLTGLQLPPTVVFQHPTPVALAERISDELAER 1023

RESULT 20

US-09-861-289-6

; Sequence 6, Application US/09861289  
 ; Patent No. US20020110897A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/861,289  
 ; CURRENT FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 11877  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces venezuelae

US-09-861-289-6

Query Match 26.7%; Score 56.5; DB 10; Length 11877;  
 Best Local Similarity 32.7%; Pred. No. 9.1e+02;  
 Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

Qy 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
 |:| | :|| |:| | :| | :|| | :|| | :| |  
 Db 11222 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVVFQHPTPVALAERISDELAER 11273

RESULT 21

US-09-860-846-6

; Sequence 6, Application US/09860846  
 ; Patent No. US20020164742A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/860,846  
 ; CURRENT FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 11877  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces venezuelae

US-09-860-846-6

Query Match 26.7%; Score 56.5; DB 10; Length 11877;  
 Best Local Similarity 32.7%; Pred. No. 9.1e+02;

Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

```
Qy      1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41
      |:| | :|| |:| | : :| | :|| | |: : : || |
Db      11222 PLREIGFDSLTAVDNRNRVRLTGLQLPPTVVFQHPTPVALAERISDELAER 11273
```

RESULT 22

US-09-836-821-6

; Sequence 6, Application US/09836821  
; Publication No. US20030087405A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/836,821  
; CURRENT FILING DATE: 2001-04-17  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 11877  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-836-821-6

Query Match 26.7%; Score 56.5; DB 11; Length 11877;  
Best Local Similarity 32.7%; Pred. No. 9.1e+02;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

```
Qy      1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41
      |:| | :|| |:| | : :| | :|| | |: : : || |
Db      11222 PLREIGFDSLTAVDNRNRVRLTGLQLPPTVVFQHPTPVALAERISDELAER 11273
```

RESULT 23

US-10-271-889-49

; Sequence 49, Application US/10271889  
; Publication No. US20030194784A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin  
; FILE REFERENCE: 600.582US1  
; CURRENT APPLICATION NUMBER: US/10/271,889  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 09/861,289  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 09/860,846  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 09/836,821



; PRIOR FILING DATE: 2001-04-17  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 49  
; LENGTH: 11877  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-10-271-889-49

Query Match 26.7%; Score 56.5; DB 12; Length 11877;  
Best Local Similarity 32.7%; Pred. No. 9.1e+02;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | :| | :|| | |:| | :| |  
Db 11222 PLREIGFDSLTAVD FRNRVNRLTGLQLPPTVV FQHPTPVALAERISDELAER 11273

#### RESULT 24

US-09-988-384B-6  
; Sequence 6, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 6  
; LENGTH: 12199  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-988-384B-6

Query Match 26.7%; Score 56.5; DB 11; Length 12199;  
Best Local Similarity 32.7%; Pred. No. 9.4e+02;  
Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
|:| | :|| |:| | :| | :|| | |:| | :| |  
Db 11544 PLREIGFDSLTAVD FRNRVNRLTGLQLPPTVV FQHPTPVALAERISDELAER 11595

#### RESULT 25

US-10-287-274-379  
; Sequence 379, Application US/10287274  
; Publication No. US20030181408A1

; GENERAL INFORMATION:  
 ; APPLICANT: Forsyth, R. Allyn  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE  
 THERETO  
 ; FILE REFERENCE: ELITRA.008DV1  
 ; CURRENT APPLICATION NUMBER: US/10/287,274  
 ; CURRENT FILING DATE: 2002-10-31  
 ; PRIOR APPLICATION NUMBER: US 60/164415  
 ; PRIOR FILING DATE: 1999-11-09  
 ; PRIOR APPLICATION NUMBER: US 09/711164  
 ; PRIOR FILING DATE: 2000-11-09  
 ; NUMBER OF SEQ ID NOS: 469  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 379  
 ; LENGTH: 196  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-10-287-274-379

Query Match 25.7%; Score 54.5; DB 12; Length 196;  
 Best Local Similarity 32.6%; Pred. No. 11;  
 Matches 15; Conservative 10; Mismatches 14; Indels 7; Gaps 2;

QY 5 ISENSLV-AMDFSGQKSR-----VIENPTEALSVAVEEGLAWRKK 43  
 | |||:| | | : :: :| :| : | : |||:| :  
 Db 109 IGENSIVGASAFVKAKAEMPANYLIVGSPAKAIRELSEQELAWKKQ 154

# RESULT 26

US-10-369-493-1220  
 ; Sequence 1220, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
 OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 1220  
 ; LENGTH: 301  
 ; TYPE: PRT  
 ; ORGANISM: Methanobacterium thermoautotrophicum  
 US-10-369-493-1220

Query Match 25.5%; Score 54; DB 12; Length 301;  
 Best Local Similarity 36.8%; Pred. No. 22;

Matches 14; Conservative 7; Mismatches 11; Indels 6; Gaps 2;

Qy 6 SENSLVAMDFSGQKSRVIENPTEA---LSVAVEEGLAW 40  
|: :|||:| | | |||| |: :|::| |  
Db 153 SQACVVAID---AKRRYIENPRESDERFIIEVDDGYCW 187

RESULT 27

US-10-369-493-22965  
; Sequence 22965, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22965  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Aeropyrum pernix  
US-10-369-493-22965

Query Match 25.5%; Score 54; DB 12; Length 432;  
Best Local Similarity 43.3%; Pred. No. 34;  
Matches 13; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 10 LVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
|: :| : |||: | | |||:| |  
Db 97 LIELDGTPNKSRLGNTTTALSIASRAAA 126

RESULT 28

US-09-769-787-9  
; Sequence 9, Application US/09769787  
; Publication No. US20030091577A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT APPLICATION NUMBER: US/09/769,787  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19

; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 234  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-9

Query Match 25.0%; Score 53; DB 11; Length 234;  
Best Local Similarity 40.0%; Pred. No. 22;  
Matches 12; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 11 VAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
||: ||: | || : ||:: || |  
Db 51 VAVFVSGKLSPAYLNPAVTIGVALKGGLPW 80

RESULT 29

US-09-738-626-4780  
; Sequence 4780, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4780  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4780

Query Match 25.0%; Score 53; DB 10; Length 530;  
Best Local Similarity 34.8%; Pred. No. 61;  
Matches 16; Conservative 10; Mismatches 14; Indels 6; Gaps 3;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVE-EGL--AWRKK 43  
| : |||||: | :||: | | :|| | |  
Db 445 PREVLDEDSLVALDAIG---AIVESVGDATSAVL DVEGLYTRWLKE 487

RESULT 30

US-10-326-671-236

; Sequence 236, Application US/10326671

; Publication No. US20030186281A1

; GENERAL INFORMATION:

; APPLICANT: Hillen, Wolfgang

; TITLE OF INVENTION: MODIFIED TETRACYCLINE REPRESSOR PROTEIN COMPOSITIONS AND METHODS OF

; TITLE OF INVENTION: USE

; FILE REFERENCE: 10182-022-999

; CURRENT APPLICATION NUMBER: US/10/326,671

; CURRENT FILING DATE: 2002-12-20

; PRIOR APPLICATION NUMBER: US 60/343,278

; PRIOR FILING DATE: 2001-12-21

; NUMBER OF SEQ ID NOS: 459

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO. 236

; LENGTH: 208

; TYPE: PRT

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: Modified tetracycline repressor

US-10-326-671-236

Query Match 24.5%; Score 52; DB 12; Length 208;

Best Local Similarity 54.2%; Pred. No. 27;

Matches 13; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 19 KSRVIENPTEALSVAVEEGLAWRK 42

||:| : | :| | | |

Db 6 KSKVINSALELLNVAGIEGLTTRK 29

RESULT 31

US-09-815-242-10691

; Sequence 10691, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 10691  
 ; LENGTH: 330  
 ; TYPE: PRT  
 ; ORGANISM: Enterococcus faecalis  
 US-09-815-242-10691

Query Match 24.5%; Score 52; DB 9; Length 330;  
 Best Local Similarity 43.8%; Pred. No. 48;  
 Matches 14; Conservative 5; Mismatches 9; Indels 4; Gaps 1;

Qy 12 AMDFSGQKSR---VIENPTEALSVAVEEGLA 39  
 ||:|:| | | | | | | | | | | |  
 Db 165 AMNFAGVKKLPVIFVVENNEYAISVPIEEQYA 196

# RESULT 32

US-10-369-493-2956

; Sequence 2956, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
 OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 2956  
 ; LENGTH: 392  
 ; TYPE: PRT  
 ; ORGANISM: Thermotoga maritima  
 US-10-369-493-2956

Query Match 24.3%; Score 51.5; DB 12; Length 392;  
 Best Local Similarity 31.8%; Pred. No. 69;  
 Matches 14; Conservative 10; Mismatches 15; Indels 5; Gaps 3;

Qy 2 MRSISEN-SLVAMDFSG---QKSRVIENPTEALSVAV-EEGLAW 40

Db 143 VKRVKESPSIVGRDLAGLVSPKEVIVENPEGDFSVVVLDSGVKW 186

RESULT 33

US-10-369-493-3798

; Sequence 3798, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3798  
; LENGTH: 617  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(617)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3798

Query Match 24.3%; Score 51.5; DB 12; Length 617;  
Best Local Similarity 35.7%; Pred. No. 1.2e+02;  
Matches 15; Conservative 9; Mismatches 15; Indels 3; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSRVIEN--PTEALSVAVEEGLAW 40  
|: || | || | : :: || ::|| ||::|  
Db 359 PVSSASILSKKKPDFIGSDIRIRDDTIPTANIAIAV-EGVSW 399

RESULT 34

US-10-369-493-11

; Sequence 11, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 11  
; LENGTH: 801  
; TYPE: PRT  
; ORGANISM: Aquifex aeolicus  
US-10-369-493-11

Query Match 24.3%; Score 51.5; DB 12; Length 801;  
Best Local Similarity 42.9%; Pred. No. 1.7e+02;  
Matches 15; Conservative 5; Mismatches 6; Indels 9; Gaps 2;

Qy 18 QKSRVIENPTE-----ALSVAV-----EEGLAWRKK 43  
:: ||| | | | | | | | | | | | | | |  
Db 171 EEGRVIELPQEMYMLIAMTLAVPEKPEERLKWAKK 205

RESULT 35

US-10-369-493-23303  
; Sequence 23303, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 23303  
; LENGTH: 335  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-10-369-493-23303

Query Match 24.1%; Score 51; DB 12; Length 335;  
Best Local Similarity 35.0%; Pred. No. 67;  
Matches 14; Conservative 8; Mismatches 10; Indels 8; Gaps 2;

Qy 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEG-----LAW 40  
|| ||: ||: ||: : : ||| | || : : |  
Db 276 SEEPLVSGDYNGNKN---SSTIDALSTMVMEGSMVKVISW 312

RESULT 36

US-10-032-585-7696  
; Sequence 7696, Application US/10032585  
; Publication No. US20030180953A1



```
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7696
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7696
```

```
Query Match          24.1%; Score 51; DB 12; Length 431;
Best Local Similarity 27.9%; Pred. No. 92;
Matches 12; Conservative 11; Mismatches 6; Indels 14; Gaps 1;
```

```
QY      1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43
          |:|:::| |          :| ||| :|| : :|::|
Db      355 PIRAVTVNS-----DNLAELQLAVNKFAYKRK 383
```

# RESULT 37

```
US-10-369-493-11560
; Sequence 11560, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11560
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11560
```

```
Query Match          23.8%; Score 50.5; DB 12; Length 406;
Best Local Similarity 41.9%; Pred. No. 1e+02;
Matches 13; Conservative 6; Mismatches 11; Indels 1; Gaps 1;
```

```
QY      8 NSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38
          | : :||| : :|:| | | | | | | |
```

RESULT 38

US-10-369-493-14587  
 ; Sequence 14587, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
 OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 14587  
 ; LENGTH: 450  
 ; TYPE: PRT  
 ; ORGANISM: Agrobacterium tumefaciens  
 US-10-369-493-14587

Query Match 23.8%; Score 50.5; DB 12; Length 450;  
 Best Local Similarity 41.9%; Pred. No. 1.1e+02;  
 Matches 13; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 8 NSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38  
 | : : || : : || | | | | :  
 Db 79 NGGICIDFS-RMNRIIEVNAEDLDVTVEPGV 108

RESULT 39

US-10-369-493-14939  
 ; Sequence 14939, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
 OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 14939

; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
US-10-369-493-14939

Query Match 23.8%; Score 50.5; DB 12; Length 465;  
Best Local Similarity 41.9%; Pred. No. 1.2e+02;  
Matches 13; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 8 NSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38  
| : : ||| : : ||| | | | | | :  
Db 94 NGGICIDFS-RMNRIIEVNAEDLDVTVEPGV 123

RESULT 40

US-10-369-493-14160

; Sequence 14160, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 14160  
; LENGTH: 487  
; TYPE: PRT  
; ORGANISM: Agrobacterium tumefaciens  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(487)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-14160

Query Match 23.8%; Score 50.5; DB 12; Length 487;  
Best Local Similarity 41.9%; Pred. No. 1.3e+02;  
Matches 13; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 8 NSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38  
| : : ||| : : ||| | | | | | :  
Db 165 NGGICIDFS-RMNRIIEVNAEDLDVTVEPGV 194

Search completed: January 13, 2004, 16:32:02  
Job time : 19.622 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:14:47 ; Search time 21.6693 Seconds  
(without alignments)  
512.073 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%  
Result Query  
No. Score Match Length DB ID Description  
-----

1	205	96.7	207	11	Q8VDI2	Q8vdi2 mus musculu
2	168	79.2	541	11	Q91WC5	Q91wc5 mus musculu
3	168	79.2	596	11	Q8BSS5	Q8bss5 mus musculu
4	168	79.2	596	11	Q8BSH4	Q8bsh4 mus musculu
5	162	76.4	447	4	Q9Y220	Q9y220 homo sapien
6	161	75.9	535	11	Q9QZC5	Q9qzc5 rattus norv
7	66	31.1	655	10	Q9C620	Q9c620 arabidopsis
8	61.5	29.0	404	16	Q8YSM7	Q8ysm7 anabaena sp
9	59	27.8	641	16	Q9ABR1	Q9abr1 caulobacter
10	58.5	27.6	602	5	Q962M6	Q962m6 plasmodium
11	58	27.4	674	2	Q9AQ18	Q9aq18 bradyrhizob
12	57.5	27.1	533	11	Q9Z2Z2	Q9z2z2 mus musculu
13	57.5	27.1	545	4	Q96JP3	Q96jp3 homo sapien
14	57.5	27.1	686	11	Q8C208	Q8c208 mus musculu
15	57	26.9	399	17	Q8PZ07	Q8pz07 methanosarc
16	56.5	26.7	1346	2	Q9ZGI2	Q9zgi2 streptomyce
17	55.5	26.2	406	5	Q9W5B5	Q9w5b5 drosophila
18	55.5	26.2	504	5	Q9NF72	Q9nf72 drosophila
19	55.5	26.2	668	3	Q06677	Q06677 saccharomyc
20	55	25.9	231	16	Q927G3	Q927g3 listeria in
21	55	25.9	732	16	Q9I1W2	Q9ilw2 pseudomonas
22	55	25.9	972	5	Q9VBX1	Q9vbx1 drosophila
23	54.5	25.7	267	17	Q8TZZ2	Q8tzz2 pyrococcus
24	54.5	25.7	491	17	Q9HM11	Q9hm11 thermoplasm
25	54.5	25.7	519	16	Q8ZCX1	Q8zcx1 yersinia pe
26	54	25.5	231	2	Q8VTU1	Q8vtu1 listeria mo
27	54	25.5	231	16	Q8Y400	Q8y400 listeria mo
28	54	25.5	242	16	Q9CLG8	Q9clg8 pasteurella
29	54	25.5	323	2	Q85118	Q85118 rhodobacter
30	54	25.5	403	16	Q8FP68	Q8fp68 corynebacte
31	54	25.5	677	2	Q8L2E8	Q8l2e8 vibrio harv
32	54	25.5	1004	17	Q8TJS3	Q8tjs3 methanosarc
33	54	25.5	1520	16	Q8D5S4	Q8d5s4 vibrio vuln
34	53	25.0	247	16	Q9A087	Q9a087 streptococc
35	53	25.0	247	16	Q8P1C7	Q8plc7 streptococc
36	53	25.0	248	16	Q8K7V7	Q8k7v7 streptococc
37	53	25.0	316	17	Q8U2R6	Q8u2r6 pyrococcus
38	53	25.0	530	16	Q8NR94	Q8nr94 corynebacte
39	53	25.0	871	4	Q96DR7	Q96dr7 homo sapien
40	52.5	24.8	211	2	Q8KNK9	Q8knk9 salmonella
41	52.5	24.8	230	16	Q8DTL0	Q8dtl0 streptococc
42	52.5	24.8	524	16	Q8ZGJ7	Q8zgj7 yersinia pe
43	52.5	24.8	590	16	Q8X2N3	Q8x2n3 escherichia
44	52.5	24.8	596	16	Q8X4F8	Q8x4f8 escherichia
45	52.5	24.8	967	16	O26050	O26050 helicobacte

#### ALIGNMENTS

##### RESULT 1

##### Q8VDI2

ID Q8VDI2 PRELIMINARY; PRT; 207 AA.  
AC Q8VDI2;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Similar to growth factor receptor-bound protein 10 (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC021820; AAH21820.1; -.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR ProDom; PD000093; SH2; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 KW Receptor.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 207 AA; 23393 MW; 02D0C5231D884882 CRC64;  
  
 Query Match 96.7%; Score 205; DB 11; Length 207;  
 Best Local Similarity 93.0%; Pred. No. 3.4e-19;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||:|||||||:||||:|||||||  
 Db 34 PMRSVSENSLVAMDFSGEKSVIDNPTEALSVAVEEGLAWRKK 76

## RESULT 2

Q91WC5

ID Q91WC5 PRELIMINARY; PRT; 541 AA.  
 AC Q91WC5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Similar to growth factor receptor bound protein 10.  
 GN GRB10.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye, and Retina;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 DR EMBL; BC016111; AAH16111.1; -.  
 DR MGD; MGI:103232; Grb10.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA\_domain.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00788; RA; 1.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR00401; SH2DOMAIN.

DR ProDom; PD000093; SH2; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00314; RA; 1.  
DR SMART; SM00252; SH2; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
DR PROSITE; PS50001; SH2; 1.  
KW Receptor.  
SQ SEQUENCE 541 AA; 61217 MW; A8FA9ED57C85F674 CRC64;

Query Match 79.2%; Score 168; DB 11; Length 541;  
Best Local Similarity 76.7%; Pred. No. 8.8e-14;  
Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
| | | | : | | | | | | | | | | | | | | : | | | | | | | | : | | | | | | | :  
Db 370 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWRKR 412

#### RESULT 3

##### Q8BSS5

ID Q8BSS5 PRELIMINARY; PRT; 596 AA.  
AC Q8BSS5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Growth factor receptor bound protein 10.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Body;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK030727; BAC27100.1; -.  
SQ SEQUENCE 596 AA; 67543 MW; EB13CA896DF41533 CRC64;

Query Match 79.2%; Score 168; DB 11; Length 596;  
Best Local Similarity 76.7%; Pred. No. 9.8e-14;  
Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
| | | | : | | | | | | | | | | | | | | : | | | | | | | | : | | | | | | | :  
Db 425 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWRKR 467

#### RESULT 4

##### Q8BSH4

ID Q8BSH4 PRELIMINARY; PRT; 596 AA.  
AC Q8BSH4;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)





Query Match 76.4%; Score 162; DB 4; Length 447;  
Best Local Similarity 74.4%; Pred. No. 4.4e-13;  
Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

## RESULT 6

```
ID Q9QZC5 PRELIMINARY; PRT; 535 AA.
AC Q9QZC5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Growth factor receptor binding protein GRB7.
GN GRB7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98421528; PubMed=9748281;
RA Kasus-Jacobi A., Perdereau D., Auzan C., Clauser E., Van Obberghen E.,
RA Mauvais-Jarvis F., Girard J., Burnol A.F.;
RT "Identification of the rat adapter Grb14 as an inhibitor of insulin
RT actions.";
RL J. Biol. Chem. 273:26026-26035(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=20260602; PubMed=10803466;
RA Kasus-Jacobi A., Bereziat V., Perdereau D., Girard J., Burnol A.F.;
RT "Evidence for an interaction between the insulin receptor and Grb7. A
RT role for two of its binding domains, PIR and SH2.";
RL Oncogene 19:2052-2059(2000).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR EMBL; AF190121; AAF01776.1; -.
DR HSSP; P35235; 1AYA.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000159; RA_domain.
DR InterPro; IPR000980; SH2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50001; SH2; 1.
```

KW Receptor.  
SQ SEQUENCE 535 AA; 59889 MW; 15DB67C4D19B89E4 CRC64;

Query Match 75.9%; Score 161; DB 11; Length 535;  
Best Local Similarity 72.1%; Pred. No. 7.4e-13;  
Matches 31; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY           1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
             | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db          366 PLRSVSDNTLVAMDFSGHAGRVIE NPQEALSAATEEAQAWRKK 408

## RESULT 7

Q9C620

```

ID Q9C620 PRELIMINARY; PRT; 655 AA.
AC Q9C620;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Receptor serine/threonine kinase PR5K, putative.
GN T4024.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC083891; AAG50590.1; -.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

```

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 655 AA; 73013 MW; 7808804B621A9566 CRC64;

Query Match 31.1%; Score 66; DB 10; Length 655;  
Best Local Similarity 30.2%; Pred. No. 4.3;  
Matches 16; Conservative 7; Mismatches 16; Indels 14; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENP-----TEALSVAVEEGLA 39  
| : || : || | || : || | : | : | : | : |  
Db 166 PSLKLEGNSFLLNDFGGSCSRNVSNPASRTALNTLESTPSTDNLKIALEDGFA 218

#### RESULT 8

##### Q8YSM7

ID Q8YSM7 PRELIMINARY; PRT; 404 AA.  
AC Q8YSM7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein Alr3057.  
GN ALR3057.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003591; BAB74756.1; -.  
DR InterPro; IPR001296; Glyco\_trans\_1.  
DR Pfam; PF00534; Glycos\_transf\_1; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 404 AA; 45485 MW; 6952C10FD5381B0B CRC64;

Query Match 29.0%; Score 61.5; DB 16; Length 404;  
Best Local Similarity 33.3%; Pred. No. 9.7;  
Matches 17; Conservative 10; Mismatches 13; Indels 11; Gaps 2;

QY 3 RSIS-----ENSLVAMDFSGQKSRVIENP--TEALSVAVEEGLAWRK 42  
||:| | || ||:: |:|: | | |: :|| : : ||:  
Db 95 RSLSSDFMHFHRLEPSLAAMNWQGEKTIFIHNDIHTQMATVADRKAILWRR 145

#### RESULT 9

##### Q9ABR1

ID Q9ABR1 PRELIMINARY; PRT; 641 AA.  
AC Q9ABR1;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein CC0165.  
 GN CC0165.  
 OS *Caulobacter crescentus*.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
 OC Caulobacteraceae; *Caulobacter*.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of *Caulobacter crescentus*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005690; AAK22152.1; -.  
 DR TIGR; CC0165; -.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR007016; Wzy\_C.  
 DR Pfam; PF04932; Wzy\_C; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 641 AA; 67175 MW; D8FF63BE76B565F9 CRC64;

Query Match 27.8%; Score 59; DB 16; Length 641;  
 Best Local Similarity 50.0%; Pred. No. 36;  
 Matches 15; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 10 LVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
 ||| | | : | ||| : || |||  
 Db 478 LVAARFGGDLALPTAPAEALASSVETGLA 507

# RESULT 10

Q962M6

ID Q962M6 PRELIMINARY; PRT; 602 AA.  
 AC Q962M6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE PV1H14040\_P.  
 GN PV1H14040C.  
 OS *Plasmodium vivax*.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; *Plasmodium*.  
 OX NCBI\_TaxID=5855;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tchavtchitch M., Fischer K., Huestis R., Saul A.;  
 RT "The sequence of 200 kb portion of a *Plasmodium vivax* chromosome  
 RT reveals a high degree of conservation with *P. falciparum* chromosome  
 RT 3.";  
 RL Mol. Biochem. Parasitol. 0:0-0(2001).  
 DR EMBL; AY003872; AAF99454.1; -.

DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 6.  
 DR PRINTS; PR00320; GPROTEINBRPT.  
 DR ProDom; PD000018; WD40; 3.  
 DR SMART; SM00320; WD40; 7.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 4.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 602 AA; 68028 MW; AFCBB6D0709AE8A4 CRC64;

Query Match 27.6%; Score 58.5; DB 5; Length 602;  
 Best Local Similarity 37.8%; Pred. No. 39;  
 Matches 14; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

Qy 7 ENSLVAMDFSGQKSRVI-ENPTEALSVAVEEGLAWRK 42  
 ||| :|| | :||:| : :|| || :| ||:  
 Db 545 ENSTLAMAFDKSESRLITTHGDKSIKVAQKKGEIWRE 581

# RESULT 11

## Q9AQ18

ID Q9AQ18 PRELIMINARY; PRT; 674 AA.  
 AC Q9AQ18;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Nodulation protein Nolo.  
 GN NOLO.  
 OS Bradyrhizobium sp. WM9.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 OX NCBI\_TaxID=133505;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WM9;  
 RA Stepkowski T., Swiderska A., Miedzinska K., Czaplinska M.,  
 RA Biesiadka J., Swiderski M., Legocki A.;  
 RT "Molecular characterization of nodulation functions, SSU rRNA and dnaK  
 RT genes in the lupin Bradyrhizobium reveals distinct phylogenetic  
 RT pathways of the symbiotic and housekeeping loci in the Bradyrhizobium  
 RT genus."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF222753; AAK00162.1; -.  
 DR InterPro; IPR003696; Carbtransf.  
 DR Pfam; PF02543; CmcH\_NodU; 1.  
 SQ SEQUENCE 674 AA; 74775 MW; 03644BA92A46C23A CRC64;

Query Match 27.4%; Score 58; DB 2; Length 674;  
 Best Local Similarity 36.4%; Pred. No. 51;  
 Matches 12; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

Qy 7 ENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
 || :: : || || || | : : : | |  
 Db 2 ENKMLCLGLSGGLDRVYENPLELPNTFLHDGAA 34

# RESULT 12

Q9Z2Z2

ID Q9Z2Z2 PRELIMINARY; PRT; 533 AA.  
AC Q9Z2Z2;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Eos protein.  
GN ZNFN1A4 OR EOS.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ICR;  
RX MEDLINE=99232954; PubMed=10218586;  
RA Homma Y., Kiyosawa H., Mori T., Oguri A., Nikaido T., Kanazawa K.,  
RA Tojo M., Takeda J., Tanno Y., Yokoya S., Kawabata I., Ikeda H.,  
RA Wanaka A.;  
RT "Eos: a novel member of the Ikaros gene family expressed predominantly  
RT in the developing nervous system.";  
RL FEBS Lett. 447:76-80(1999).  
DR EMBL; AB017615; BAA36213.1; -.  
DR HSSP; P15822; 1BBO.  
DR MGD; MGI:1343139; Znf1a4.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 6.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00355; Znf\_C2H2; 6.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 533 AA; 58167 MW; 7A5FF32C6FFDC372 CRC64;

Query Match 27.1%; Score 57.5; DB 11; Length 533;  
Best Local Similarity 36.4%; Pred. No. 46;  
Matches 16; Conservative 8; Mismatches 15; Indels 5; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 39  
| ||:| ||: :| ::| : || | || ||: |:  
Db 37 PSRSLSANSIKVMYSDEESSRLLGPDERLLDKDDSVIVEDSLS 80

# RESULT 13

Q96JP3

ID Q96JP3 PRELIMINARY; PRT; 545 AA.  
AC Q96JP3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein KIAA1782 (Fragment).  
GN KIAA1782.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=21245130; PubMed=11347906;  
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large Proteins in vitro.";  
 RL DNA Res. 8:85-95(2001).  
 DR EMBL; AB058685; BAB47411.1; -.  
 DR Genew; HGNC:13179; ZNFN1A4.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 5.  
 DR ProDom; PD000003; Znf\_C2H2; 1.  
 DR SMART; SM00355; Znf\_C2H2; 6.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 4.  
 KW Hypothetical protein; Metal-binding; Zinc; Zinc-finger.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 545 AA; 59742 MW; 7A8539E5B8F9BD84 CRC64;

Query Match 27.1%; Score 57.5; DB 4; Length 545;  
 Best Local Similarity 36.4%; Pred. No. 47;  
 Matches 16; Conservative 8; Mismatches 15; Indels 5; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 39  
 | ||:| ||: :| ::| :| || || ||:| :  
 Db 50 PSRSLANSIKVEMYSDDEESSRLLGPDERLLEKDDSVIVEDSLS 93

#### RESULT 14

##### Q8C208

ID Q8C208 PRELIMINARY; PRT; 686 AA.  
 AC Q8C208;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Zinc finger protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK089522; BAC40912.1; -.  
 SQ SEQUENCE 686 AA; 75078 MW; F99ADB635184FAC0 CRC64;

Query Match 27.1%; Score 57.5; DB 11; Length 686;  
 Best Local Similarity 36.4%; Pred. No. 61;  
 Matches 16; Conservative 8; Mismatches 15; Indels 5; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEAL-----SVAVEEGLA 39  
 | | | : | | | : | : | : | | | | | : | :  
 Db 90 PSRSLSANSIKVEMYSDDESSRLGPDRLLDKDDSVIVEDSLS 133

RESULT 15

Q8PZ07

ID Q8PZ07 PRELIMINARY; PRT; 399 AA.  
 AC Q8PZ07;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Conserved protein.  
 GN MM0691.  
 OS Methanosarcina mazei (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,  
 RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,  
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazei: evidence for lateral gene  
 RT transfer between Bacteria and Archaea.";  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AE013294; AAM30387.1; -.  
 DR InterPro; IPR003806; DUF201.  
 DR Pfam; PF02655; DUF201; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 399 AA; 43608 MW; 5ECDD65F360A1B9D CRC64;

Query Match 26.9%; Score 57; DB 17; Length 399;  
 Best Local Similarity 37.0%; Pred. No. 38;  
 Matches 17; Conservative 7; Mismatches 16; Indels 6; Gaps 3;

QY 1 PMRSISENSLVAMDF-SGQKS--RVIENPTEALSVAVEE---GLAW 40  
 | : :: | | : | | | | : | | : | | | : |  
 Db 170 PDQGLTENEIVIQFLEGTPSSSVSVLSTKDEALAVAVNEQLTGIPW 215.

RESULT 16

Q9ZGI2

ID Q9ZGI2 PRELIMINARY; PRT; 1346 AA.  
 AC Q9ZGI2;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Type I polyketide synthase PikaIV.  
 GN PIKAIV.  
 OS Streptomyces venezuelae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;



OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=54571;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC15439;  
 RX MEDLINE=98445333; PubMed=9770448;  
 RA Xue Y., Zhao L., Liu H.w., Sherman D.H.;  
 RT "A gene cluster for macrolide antibiotic biosynthesis in streptomyces  
 RT venezuelae: architecture of metabolic diversity.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).  
 DR EMBL; AF079138; AAC69332.1; -.  
 DR HSSP; P25715; 1MLA.  
 DR InterPro; IPR001227; Ac\_transferase.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR000734; Lipase.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR InterPro; IPR001031; Thioesterase.  
 DR Pfam; PF00698; Acyl\_transf; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 1.  
 DR Pfam; PF00550; pp-binding; 1.  
 DR Pfam; PF00975; Thioesterase; 1.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 KW Phosphopantetheine; Transferase.  
 SQ SEQUENCE 1346 AA; 141913 MW; 3E149C8044FBE5F2 CRC64;

Query Match 26.7%; Score 56.5; DB 2; Length 1346;  
 Best Local Similarity 32.7%; Pred. No. 1.8e+02;  
 Matches 17; Conservative 10; Mismatches 14; Indels 11; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSR-----VIENPTE-ALSVAVEEGLAWR 41  
 | : | : | | : | | : : | | : : | | : : | |  
 Db 972 PLREIGFDSLTAVDFRNRVNRLTGLQLPPTVVVFQHPTPVALAERISDELAER 1023

# RESULT 17

Q9W5B5

ID Q9W5B5 PRELIMINARY; PRT; 406 AA.  
 AC Q9W5B5;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG14630 protein.  
 GN EG:BACR7A4.9 OR CG14630.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003419; AAF45580.2; -.  
 DR FlyBase; FBgn0014903; EG:BACR7A4.9.  
 DR InterPro; IPR004994; Gamma-BBH.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF03322; Gamma-BBH; 1.  
 DR PROSITE; PS00038; HLH\_1; 1.  
 SQ SEQUENCE 406 AA; 46910 MW; 87C93B72AD1DB1D2 CRC64;  
  
 Query Match 26.2%; Score 55.5; DB 5; Length 406;  
 Best Local Similarity 24.4%; Pred. No. 62;  
 Matches 11; Conservative 14; Mismatches 17; Indels 3; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTE---ALSVAVEEGLAWRK 42  
 | : : : | : : : | : : : : : | : : : : : |  
 Db 290 PFHSLWRAPVICLDVDGRFARINQNTTKRDSRFSVSLAQAVSWYK 334

# RESULT 18

Q9NF72

ID Q9NF72 PRELIMINARY; PRT; 504 AA.  
 AC Q9NF72;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE EG:BACR7A4.9 protein.  
 GN EG:BACR7A4.9 OR CG14630.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Papagiannakis G., Spanos L., Bolshakov V., Siden-Kiamos I., Louis C.;  
 RT "Sequencing the distal X chromosome of Drosophila melanogaster.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Benos P.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL109630; CAB51679.1; -.

DR FlyBase; FBgn0014903; EG:BACR7A4.9.  
 DR InterPro; IPR004994; Gamma-BBH.  
 DR Pfam; PF03322; Gamma-BBH; 1.  
 SQ SEQUENCE 504 AA; 57985 MW; 1E4DCEA775BB167E CRC64;

Query Match 26.2%; Score 55.5; DB 5; Length 504;  
 Best Local Similarity 24.4%; Pred. No. 79;  
 Matches 11; Conservative 14; Mismatches 17; Indels 3; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTE---ALSVAVEEGLAWRK 42  
 | | : : : | | : : : | | : : : | | : : : | |  
 Db 388 PFHSLWRAPVICLDVDGRFARINQNTTKRDSRFSVSLAQAVSWYK 432

# RESULT 19

Q06677

ID Q06677 PRELIMINARY; PRT; 668 AA.  
 AC Q06677;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE SIMILARITY to human transformation-sensitive protein IEF.  
 GN SWA2 OR D9798.10 OR YDR320C.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
 RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
 RA Taich A., Trevaskis E., Vignati D., Wilcox L., Wohldman P., Vaudin M.,  
 RA Wilson R., Waterston R.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Du Z.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Waterston R.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Jia Y., Cherry J.M.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; U32517; AAB64756.1; -.  
 DR SGD; S0002728; SWA2.  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 3.

SQ SEQUENCE 668 AA; 75019 MW; CCDF1F78315E3D44 CRC64;

Query Match 26.2%; Score 55.5; DB 3; Length 668;  
Best Local Similarity 29.5%; Pred. No. 1.1e+02;  
Matches 13; Conservative 11; Mismatches 19; Indels 1; Gaps 1;

QY 1 PMRSISENSLVAMDFS-GQKSRVIENPTEALSVAVEEGLAWRKK 43  
|:| |: ::::| |: |: ||| : || : |: |  
Db 409 PLRIIALSNIIASQLKIGEYSKIENSSMALELFPSSKAKWKNK 452

RESULT 20

Q927G3

ID Q927G3 PRELIMINARY; PRT; 231 AA.  
AC Q927G3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein lin2826.  
GN LIN2826.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / Serovar 6a;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";   
RL Science 294:849-852(2001).  
DR EMBL; AL596173; CAC98052.1; -.  
DR ListiList; LIN02826; -.  
DR InterPro; IPR001789; Response\_reg.  
DR InterPro; IPR001867; Trans\_reg\_C.  
DR Pfam; PF00072; response\_reg; 1.  
DR Pfam; PF00486; trans\_reg\_C; 1.  
DR ProDom; PD000039; Response\_reg; 1.  
DR ProDom; PD000329; Trans\_reg\_C; 1.  
DR SMART; SM00448; REC; 1.  
DR PROSITE; PS50110; RESPONSE\_REGULATORY; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 231 AA; 26090 MW; 2AE7B9F01967A8B3 CRC64;

Query Match 25.9%; Score 55; DB 16; Length 231;  
Best Local Similarity 27.8%; Pred. No. 37;  
Matches 10; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
||| : :: | : :: : || | : | | : : |  
Db 192 SENQALRVNMSNIRRKIEQNPAEPAYILTEVGVGYS 227

RESULT 21

Q9I1W2

ID Q9I1W2 PRELIMINARY; PRT; 732 AA.

AC Q9I1W2;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE 1,4-alpha-glucan branching enzyme.

GN GLGB OR PA2153.

OS *Pseudomonas aeruginosa*.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.

RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RE "Catalytic Activity of the *Escherichia coli*  $\beta$ -Glucuronidase

RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
RT *opportunistic pathogen*

RT opportunistic pathogen.";  
BI *Nature* 406:250-261 (2003).

RL Nature 406:959-964 (2000).  
DB EMBL AF004648; JG3774

DR EMBL; AE004642; AAG05541.1; -.

DR InterPro; IPR006047; Alpha\_amyl\_cat.

DR InterPro; IPR006407; GlgB.

DR InterPro; IPR004193; Glyco\_hydro\_13N.

DR InterPro; IPR001484; Pyrokinin.  
 DE Pfam; PF02122; Pyrokinin.

DR Pfam; PF00128; alpha-amylase; 1.

DR Pfam; PF02922; isoamylase\_N; 2.  
DR TIGRFAMs; TIGR01515; branching\_enzym; 1.

DR PROSITE; PS00539; PYROKININ; 1.

KW Complete proteome.

SQ SEQUENCE 732 AA; 82562 MW; C31303D6D5F929F4 CRC64;

Query Match 25.9%; Score 55; DB 16; Length 732;

Best Local Similarity 32.5%; Pred. No. 1.4e+02;

Matches 13; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40

Db 431 PNRHGGRENLEAIDFLHHLNQVVAETPGALVIAEESTAW 470

RESULT 22

Q9VBX1

ID Q9VBX1 PRELIMINARY; PRT; 972 AA.

AC Q9VBX1;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE CG11847 protein.  
 GN CG11847.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003750; AAF56406.1; -.  
 DR FlyBase; FBgn0039281; CG11847.  
 SQ SEQUENCE 972 AA; 110214 MW; A06FF57ECADEF9C3 CRC64;

Query Match 25.9%; Score 55; DB 5; Length 972;  
 Best Local Similarity 25.4%; Pred. No. 2e+02;

Matches 16; Conservative 8; Mismatches 17; Indels 22; Gaps 2;

```
QY      3 RSISENSLVA-----MDFSGQKSRVIENPT-----EALSVAVEEGLAW 40
      |  :| |:          : | | :|||          || |:|: :||
Db      524 RDAQQNELIVKRYMRPKDIYVHAEIQGASSVIIQNPTGEEIPPKTLLEAGSMAISYSVAW 583

QY      41 RKK 43
      |
Db      584 DAK 586
```

#### RESULT 23

Q8TZZ2

ID Q8TZZ2 PRELIMINARY; PRT; 267 AA.  
AC Q8TZZ2;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Regulatory protein.  
GN PF1832.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE010279; AAL81956.1; -.  
DR InterPro; IPR003801; DUF198.  
DR Pfam; PF02649; DUF198; 1.  
KW Complete proteome.  
SQ SEQUENCE 267 AA; 30509 MW; 47AA23BCEF2DE7BE CRC64;

Query Match 25.7%; Score 54.5; DB 17; Length 267;  
Best Local Similarity 45.2%; Pred. No. 52;  
Matches 14; Conservative 6; Mismatches 6; Indels 5; Gaps 1;

```
QY      11 VAMDFSGQK-----SRVIENPTEALSVAVEE 36
      ||:| :| ||::|: || :| ||||
Db      45 VAIDLPEEKKGIHMSRLVESITETMSEAVEE 75
```

#### RESULT 24

Q9HM11

ID Q9HM11 PRELIMINARY; PRT; 491 AA.  
AC Q9HM11;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE PurH bifunctional enzyme related protein.  
GN TA0060.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
OC Thermoplasmataceae; Thermoplasma.



OX NCBI\_TaxID=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM.1728;  
 RX MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
 RT acidophilum.";  
 RL Nature 407:508-513(2000).  
 DR EMBL; AL445063; CAC11208.1; -.  
 DR InterPro; IPR002695; AICARFT\_IMPCHas.  
 DR InterPro; IPR001179; FKBP\_PPIase.  
 DR InterPro; IPR004362; MGS\_like.  
 DR Pfam; PF01808; AICARFT\_IMPCHas; 1.  
 DR Pfam; PF02142; MGS; 1.  
 DR ProDom; PD004666; AICARFT\_IMPCHas; 1.  
 DR PROSITE; PS00453; FKBP\_PPIASE\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 491 AA; 54075 MW; 8F2A231DCA0B7FCC CRC64;

Query Match 25.7%; Score 54.5; DB 17; Length 491;  
 Best Local Similarity 32.6%; Pred. No. 1e+02;  
 Matches 14; Conservative 9; Mismatches 11; Indels 9; Gaps 1;

QY 5 ISENSLVAMDFSGQKSRVIENPTEA-----LSVAVEEGL 38  
 :: :| ||: ::|:| | ||| :| : || | |  
 Db 181 LASDSYVAIGYNGEKLRYGENPDQAGYLFSDPSVGVAASEKL 223

# RESULT 25

## Q8ZCX1

ID Q8ZCX1 PRELIMINARY; PRT; 519 AA.  
 AC Q8ZCX1;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative exopolyphosphatase (EC 3.6.1.11).  
 GN PPX OR YPO2837 OR Y1397.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RL Nature 413:523-527(2001).  
 RN [2]



SQ SEQUENCE 231 AA; 26093 MW; B1054F124CD931C3 CRC64;

Query Match 25.5%; Score 54; DB 2; Length 231;  
Best Local Similarity 27.8%; Pred. No. 51;  
Matches 10; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
||| : :: | : :: : || | : | : : |  
Db 192 SENQALRVNMSNIRRKIEKNPAEPAYILTEVGVGGR 227

RESULT 27

Q8Y400

ID Q8Y400 PRELIMINARY; PRT; 231 AA.  
AC Q8Y400;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo2678.  
GN LMO2678.  
OS *Listeria monocytogenes*.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-e / Serovar 1/2a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of *Listeria* species.";  
RL Science 294:849-852(2001).  
DR EMBL; AL591984; CAD00891.1; -.  
DR ListiList; LMO02678; -.  
DR InterPro; IPR001789; Response\_reg.  
DR InterPro; IPR001867; Trans\_reg\_C.  
DR Pfam; PF00072; response\_reg; 1.  
DR Pfam; PF00486; trans\_reg\_C; 1.  
DR ProDom; PD000039; Response\_reg; 1.  
DR ProDom; PD000329; Trans\_reg\_C; 1.  
DR SMART; SM00448; REC; 1.  
DR PROSITE; PS50110; RESPONSE\_REGULATORY; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 231 AA; 26178 MW; 29F8F9C92171D245 CRC64;  
  
Query Match 25.5%; Score 54; DB 16; Length 231;  
Best Local Similarity 27.8%; Pred. No. 51;  
Matches 10; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 ||| : :: | : :: || | : | : : |  
 Db 192 SENQALRVNMSNIRRKIEKNPAEPAYILTEVGVGYS 227

# RESULT 28

Q9CLG8

ID Q9CLG8 PRELIMINARY; PRT; 242 AA.  
 AC Q9CLG8;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein PM1266.  
 GN PM1266.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 DR EMBL; AE006165; AAK03350.1; -.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 242 AA; 27098 MW; 17D821923C156EE6 CRC64;

Query Match 25.5%; Score 54; DB 16; Length 242;  
 Best Local Similarity 28.6%; Pred. No. 54;  
 Matches 12; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRK 42  
 | | : : | : : | : : : | : | | : | |  
 Db 81 PWLSVLDNVQLHLHLQGGKKNQSEEKAKALLTAVKMASHWHK 122

# RESULT 29

O85118

ID O85118 PRELIMINARY; PRT; 323 AA.  
 AC O85118;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Flagellar switch protein.  
 GN FLIM.  
 OS Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
 OC Rhodobacteraceae; Rhodobacter.  
 OX NCBI\_TaxID=1063;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WS8;  
 RX MEDLINE=98348462; PubMed=9683497;  
 RA Garcia N., Campos A., Osorio A., Poggio S., Gonzalez-Pedrajo B.,  
 RA Camarena L., Dreyfus G.;  
 RT "The flagellar switch genes *fliM* and *fliN* of *Rhodobacter sphaeroides*  
 RT are contained in a large flagellar gene cluster.";  
 RL J. Bacteriol. 180:3978-3982(1998).  
 DR EMBL; AF044254; AAC32319.1; -..  
 DR InterPro; IPR001689; Flag\_FliM.  
 DR InterPro; IPR001543; SpoA.  
 DR Pfam; PF02154; FliM; 1.  
 DR Pfam; PF01052; SpoA; 1..  
 DR ProDom; PD001777; SpoA; 1.  
 DR TIGRFAMs; TIGR01397; *fliM\_switch*; 1.  
 SQ SEQUENCE 323 AA; 36502 MW; EE5649D23165526A CRC64;

Query Match 25.5%; Score 54; DB 2; Length 323;  
 Best Local Similarity 42.9%; Pred. No. 75;  
 Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

QY 14 DFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 :|: : ||| |: |:|:: :|||  
 Db 144 EFTATEERVIELVTDRLNVALQ--VAWR 169

#### RESULT 30

##### Q8FP68

ID Q8FP68 PRELIMINARY; PRT; 403 AA.  
 AC Q8FP68;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative DNA processing protein.  
 GN CE1918.  
 OS *Corynebacterium efficiens*.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
 RA Usuda Y., Sugimoto S.;  
 RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005220; BAC18728.1; -..  
 KW Complete proteome.  
 SQ SEQUENCE 403 AA; 43219 MW; DCB5A2A6C419EAF0 CRC64;

Query Match 25.5%; Score 54; DB 16; Length 403;  
 Best Local Similarity 36.8%; Pred. No. 97;  
 Matches 14; Conservative 5; Mismatches 13; Indels 6; Gaps 1;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGL 38

Db 330 PIQGLSRNELRVYDALGR-----HPREAAEVATETGL 361

RESULT 31

Q8L2E8

ID Q8L2E8 PRELIMINARY; PRT; 677 AA.  
 AC Q8L2E8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Zinc metalloprotease Pap6.  
 OS *Vibrio harveyi*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=669;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Teo J., Poh C.L., Zhang L.H.;  
 RT "Vibrio harveyi zinc metalloprotease."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF508306; AAM34261.1; -.  
 DR InterPro; IPR001570; Peptidase\_M4.  
 DR InterPro; IPR005075; Pep\_M4\_propep.  
 DR InterPro; IPR006025; Zn\_MTpeptdse.  
 DR Pfam; PF01447; Peptidase\_M4; 1.  
 DR Pfam; PF02868; Peptidase\_M4\_C; 1.  
 DR Pfam; PF03413; Pep\_M4\_propep; 1.  
 DR PRINTS; PR00730; THERMOLYSIN.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Protease; Metalloprotease.  
 SQ SEQUENCE 677 AA; 75120 MW; 5E904C0A127CA186 CRC64;

Query Match 25.5%; Score 54; DB 2; Length 677;  
 Best Local Similarity 34.2%; Pred. No. 1.8e+02;  
 Matches 13; Conservative 12; Mismatches 9; Indels 4; Gaps 1;

QY 2 MRSISENSLVAMDFSGQKSRVIENP---TEALSVAVE 35  
 ||::: ||| |:::|: :|||::| |  
 Db 1 MRNVTLLSLVPFAFASQAAQIVEHSQTDLSEALNIAGE 38

RESULT 32

Q8TJS3

ID Q8TJS3 PRELIMINARY; PRT; 1004 AA.  
 AC Q8TJS3;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE TPR-domain containing protein.  
 GN MA3704.  
 OS *Methanosarcina acetivorans*.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; *Methanosarcina*.  
 OX NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.;  
 RT "The genome of *Methanosarcina acetivorans* reveals extensive metabolic  
 RT and physiological diversity.";  
 RL Genome Res. 12:532-542(2002).  
 DR EMBL; AE011082; AAM07059.1; -.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 19.  
 DR SMART; SM00028; TPR; 18.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 1004 AA; 112398 MW; 51B5D3F7A777DD3D CRC64;

Query Match 25.5%; Score 54; DB 17; Length 1004;  
 Best Local Similarity 32.6%; Pred. No. 2.8e+02;  
 Matches 14; Conservative 5; Mismatches 18; Indels 6; Gaps 1;

QY 7 ENSLVAMDFS-----GQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||| : | || : | : || : : | || |  
 Db 335 ENSCIMSGIGEIIYQLGDYSRALEAFEQALRLDIENGFAWNGK 377

# RESULT 33

Q8D5S4

ID Q8D5S4 PRELIMINARY; PRT; 1520 AA.  
 AC Q8D5S4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Non-ribosomal peptide synthetase modules.  
 GN VV20831.  
 OS *Vibrio vulnificus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE016810; AAO07755.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 1520 AA; 169111 MW; A07B82C327F9BCE6 CRC64;

Query Match 25.5%; Score 54; DB 16; Length 1520;  
Best Local Similarity 37.8%; Pred. No. 4.5e+02;  
Matches 14; Conservative 8; Mismatches 11; Indels 4; Gaps 2;

QY 9 SLVAMDFSG-QKSRVIEN---PTEALSVAVEEGLAWR 41  
:| :||| :|:| | | | : ||:|  
Db 798 ALEHLDFSGVDVNRLLMNGSSPALALPVVITNGLSWQ 834

RESULT 34

Q9A087

ID Q9A087 PRELIMINARY; PRT; 247 AA.  
AC Q9A087;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical protein SPY0887.  
GN SPY0887.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL; AE006538; AAK33807.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 247 AA; 27886 MW; 92F24F4F6A62A5DF CRC64;

Query Match 25.0%; Score 53; DB 16; Length 247;  
Best Local Similarity 35.7%; Pred. No. 75;  
Matches 10; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 2 MRSISENSLVAMDFSGQKSRVIENPTEA 29  
:::: || || | | :: :: || |  
Db 120 LKTLKENHLVVGDLSSKERQIIENSMPI 147

RESULT 35

Q8P1C7

ID Q8P1C7 PRELIMINARY; PRT; 247 AA.  
AC Q8P1C7;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein spyM18\_0948.  
GN SPYM18\_0948.  
OS Streptococcus pyogenes (serotype M18).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;



OC Streptococcus.  
 OX NCBI\_TaxID=186103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS8232 / Serotype M18;  
 RX MEDLINE=21927593; PubMed=11917108;  
 RA Smoot J.C., Barbian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 DR EMBL; AE010023; AAL97590.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 247 AA; 27688 MW; 8128E5E5CB73B4CE CRC64;

Query Match 25.0%; Score 53; DB 16; Length 247;  
 Best Local Similarity 35.7%; Pred. No. 75;  
 Matches 10; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 2 MRSISENSLVAMDFSGQKSRVIENPTEA 29  
 :::: || || | | :: ::||| |  
 Db 120 LKTLKENHLVVGDLSSKERQIENSMPA 147

# RESULT 36

Q8K7V7

ID Q8K7V7 PRELIMINARY; PRT; 248 AA.  
 AC Q8K7V7;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein SpyM3\_0606.  
 GN SPYM3\_0606.  
 OS Streptococcus pyogenes (serotype M3).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=198466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS315 / Serotype M3;  
 RX MEDLINE=22133808; PubMed=12122206;  
 RA Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,  
 RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
 RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
 RA Schlievert P.M., Musser J.M.;  
 RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
 RT phage-encoded toxins, the high-virulence phenotype, and clone  
 RT emergence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
 DR EMBL; AE014149; AAM79213.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 248 AA; 27862 MW; CD73A3F3606B73B4 CRC64;

Query Match 25.0%; Score 53; DB 16; Length 248;

Best Local Similarity 35.7%; Pred. No. 75;  
Matches 10; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

Qy 2 MRSISENSLVAMDFSGQKSRVIENPTEA 29  
::: || || | | :: ::|| |  
Db 120 LKTLKENHLVVVDLSSKERQIIENSMIPA 147

RESULT 37

Q8U2R6

ID Q8U2R6 PRELIMINARY; PRT; 316 AA.  
AC Q8U2R6;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative dehydrogenase.  
GN PF0766.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE010194; AAL80890.1; -.  
DR InterPro; IPR000683; GFO\_IDH\_MocA.  
DR InterPro; IPR004104; GFO\_IDH\_MocA\_C.  
DR Pfam; PF01408; GFO\_IDH\_MocA; 1.  
DR Pfam; PF02894; GFO\_IDH\_MocA\_C; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 316 AA; 35432 MW; 5C0359EE24A76B2E CRC64;

Query Match 25.0%; Score 53; DB 17; Length 316;  
Best Local Similarity 28.6%; Pred. No. 99;  
Matches 12; Conservative 11; Mismatches 13; Indels 6; Gaps 1;

QY 5 ISENSLVAMDFSGQKSRVIE-----NPTEALSVAVEEGLAW 40  
: ::|: : || || :|| : | |: ||:|:  
Db 201 VEDHALIMLGFSNGKSGIETNWLTPHKTRTLTAVGTEGIAY 242

RESULT 38

Q8NR94

ID Q8NR94 PRELIMINARY; PRT; 530 AA.  
AC Q8NR94;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent  
DE oxidoreductases (EC 1.14.13.50).  
GN CGL1158.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.

OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AP005277; BAB98551.1; -.  
 DR InterPro; IPR000733; Flav\_monooxygenase.  
 DR InterPro; IPR002938; Moxy\_FAD\_binding.  
 DR InterPro; IPR003042; Rng\_mnoxygenase.  
 DR Pfam; PF01494; FAD\_binding\_3; 1.  
 DR Pfam; PF01360; Monooxygenase; 1.  
 DR PRINTS; PR00420; RINGMNOXGNASE.  
 KW Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 530 AA; 59195 MW; D989081DCC50B1F6 CRC64;

Query Match 25.0%; Score 53; DB 16; Length 530;  
 Best Local Similarity 34.8%; Pred. No. 1.8e+02;  
 Matches 16; Conservative 10; Mismatches 14; Indels 6; Gaps 3;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVE-EGL--AWRKK 43  
 | : |||||: | ::|: :| | :: || | | :  
 Db 445 PREVLDESLVALDAIG--AIVESVGDATSAVL DVEGLYTRWLKE 487

# RESULT 39

Q96DR7

ID Q96DR7 PRELIMINARY; PRT; 871 AA.  
 AC Q96DR7;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative SH3 domain-containing guanine exchange factor SGEF.  
 GN SGEF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostatic carcinoma;  
 RA Qi H., Fillion C., Labrie Y., Grenier J., Fournier A., Labrie C.;  
 RT "Isolation and androgen regulation of human CSGEF, a splicing variant  
 of a new putative member (SGEF) of Dbl family, that maps to 3q25.31.";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; AF415175; AAL27001.1; -.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000219; RhoGEF.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00621; RhoGEF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00233; PH; 1.

DR SMART; SM00325; RhoGEF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50010; DH\_2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW SH3 domain.  
 SQ SEQUENCE 871 AA; 97402 MW; 326080B5A2999F60 CRC64;

Query Match 25.0%; Score 53; DB 4; Length 871;  
 Best Local Similarity 35.5%; Pred. No. 3.2e+02;  
 Matches 11; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALS 31  
 | : | | | | | : : : | | : | |  
 Db 206 PQKSSSEQKLPLQRLPSQENELLENPSVVL 236

# RESULT 40

## Q8KNK9

ID Q8KNK9 PRELIMINARY; PRT; 211 AA.  
 AC Q8KNK9;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE TraW.  
 GN TRAW.  
 OS Salmonella typhi.  
 OG Plasmid pED208.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87056998; PubMed=2877970;  
 RA Finlay B.B., Frost L.S., Paranchych W.;  
 RT "Nucleotide sequence of the tra YALE region from IncFV plasmid  
 pED208.";  
 RL J. Bacteriol. 168:990-998(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92048497; PubMed=1943709;  
 RA Di Laurenzio L., Frost L.S., Finlay B.B., Paranchych W.;  
 RT "Characterization of the oriT region of the IncFV plasmid pED208.";  
 RL Mol. Microbiol. 5:1779-1790(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22195890; PubMed=12206753;  
 RA Lu J., Manchak J., Klimke W., Davidson C., Firth N., Skurray R.A.,  
 RA Frost L.S.;  
 RT "Analysis and Characterization of IncFV Plasmid pED208 Transfer  
 Region.";  
 RL Plasmid 48:24-37(2002).  
 DR EMBL; AF411480; AAM90715.1; -.  
 DR InterPro; IPR001179; FKBP\_PPIase.  
 DR PROSITE; PS00453; FKBP\_PPIASE\_1; 1.  
 KW Plasmid.  
 SQ SEQUENCE 211 AA; 23812 MW; 5E3C37E2F17BF0D2 CRC64;

Query Match 24.8%; Score 52.5; DB 2; Length 211;  
Best Local Similarity 36.2%; Pred. No. 72;  
Matches 17; Conservative 6; Mismatches 11; Indels 13; Gaps 2;

```
Qy      2 MRSISENSLVAMDFSGQ-----KSRVIEN-----PTEALSVAVE 35
      | : : | ||: ||: | |||| | | |::| |
Db      34 MLTTIQTRLKAMEASGEMAREQEAFKQRVIENTLRPRPVEGLTLAQE 80
```

Search completed: January 13, 2004, 16:22:11  
Job time : 24.6693 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 16:17:58 ; Search time 7.11024 Seconds  
(without alignments)  
284.400 Million cell updates/sec

Title: US-09-936-697-5  
Perfect score: 212  
Sequence: 1 PMRSISENSLVAMDFSGQKS.....ENPTEALSVAVEEGLAWRKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		% Match	Length			
1	212	100.0	540	1	GRBE_HUMAN	Q14449 homo sapien
2	205	96.7	538	1	GRBE_MOUSE	Q9jlm9 mus musculu
3	205	96.7	538	1	GRBE_RAT	O88900 rattus norv
4	169	79.7	594	1	GRBA_HUMAN	Q13322 homo sapien
5	162	76.4	532	1	GRB7_HUMAN	Q14451 homo sapien
6	161	75.9	621	1	GRBA_MOUSE	Q60760 mus musculu
7	159	75.0	535	1	GRB7_MOUSE	Q03160 mus musculu
8	58.5	27.6	685	1	YGO4_YEAST	P53118 saccharomyc
9	54.5	25.7	196	1	PAAY_ECOLI	P77181 escherichia
10	54	25.5	274	1	HIS6_METTH	O27398 methanobact
11	54	25.5	416	1	ENO_SULTO	Q972b6 sulfolobus
12	54	25.5	432	1	ENO_AERPE	Q9y927 aeropyrum p
13	54	25.5	513	1	GUAA_BACHD	Q9kf78 bacillus ha
14	53	25.0	579	1	DLDi_KLULA	Q12627 kluyveromyc
15	52	24.5	336	1	NADA_HELPY	O25910 helicobacte
16	52	24.5	447	1	YPEB_OCEIH	P59106 oceanobacil
17	52	24.5	472	1	6PGD_LACLC	P96789 lactococcus

18	51.5	24.3	185	1	NUSG_TREPA	083264	treponema p
19	51.5	24.3	392	1	CARA_THEMA	Q9wz28	thermotoga
20	51.5	24.3	476	1	MPPB_NEUCR	P11913	neurospora
21	51.5	24.3	801	1	RIR1_AQUAE	O66503	aquifex aeo
22	51.5	24.3	814	1	OPHL_HUMAN	Q9unal	homo sapien
23	51	24.1	234	1	GLPF_STRPN	P52281	streptococc
24	51	24.1	334	1	G3P1_BACSU	P09124	bacillus su
25	51	24.1	475	1	TPS1_PICAN	O94213	pichia angu
26	50.5	23.8	192	1	BM3R_BACME	P43506	bacillus me
27	50.5	23.8	593	1	VG13_BPML5	Q05219	mycobacteri
28	50.5	23.8	595	1	VG13_BPMD2	O64206	mycobacteri
29	50.5	23.8	678	1	ABG1_HUMAN	P45844	homo sapien
30	50.5	23.8	993	1	YIS2_YEAST	P40562	saccharomyc
31	50.5	23.8	2109	1	RRPL_VSVJO	P16379	vesicular s
32	50	23.6	336	1	NADA_HELPJ	Q9zjn1	helicobacte
33	50	23.6	376	1	NIV2_ANASP	P58637	anabaena sp
34	49.5	23.3	672	1	GYS_CAEEL	Q9u2d9	caenorhabdi
35	49.5	23.3	693	1	LYS4_YEAST	P49367	saccharomyc
36	49	23.1	104	1	Y4EB_RHISN	P55425	rhizobium s
37	49	23.1	461	1	GATB_METKA	Q8tws2	methanopyru
38	49	23.1	557	1	HLYB_SERMA	P15321	serratia ma
39	49	23.1	602	1	PEX5_HUMAN	P50542	homo sapien
40	49	23.1	639	1	PEX5_MOUSE	O09012	mus musculu
41	49	23.1	662	1	GARP_HUMAN	Q14392	homo sapien
42	49	23.1	896	1	TPS2_YEAST	P31688	saccharomyc
43	48.5	22.9	573	1	ILVI_HAEIN	P45261	haemophilus
44	48.5	22.9	576	1	MUTL_CHLMU	Q9pjpg5	chlamydia m
45	48.5	22.9	666	1	ABG1_MOUSE	Q64343	mus musculu

#### ALIGNMENTS

##### RESULT 1

##### GRBE\_HUMAN

ID GRBE\_HUMAN STANDARD; PRT; 540 AA.  
AC Q14449;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Growth factor receptor-bound protein 14 (GRB14 adapter protein).  
GN GRB14.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96218175; PubMed=8647858;  
RA Daly R.J., Sanderson G.M., Janes P.W., Sutherland R.L.;  
RT "Cloning and characterization of GRB14, a novel member of the GRB7  
RT gene family.";  
RL J. Biol. Chem. 271:12502-12510(1996).  
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE  
CC AUTOPHOSPHORYLATED INSULIN RECEPTOR WHICH IS THEN INHIBITED. THE  
CC INTERACTION IS MEDIATED BY THE SH2 DOMAIN (BY SIMILARITY).  
CC -!- SUBUNIT: Binds to the ankyrin repeat region of TNKS2 via its N-

CC terminus.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and  
 CC endosomes.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE LIVER, KIDNEY,  
 CC PANCREAS, TESTIS, OVARY, HEART, AND SKELETAL MUSCLE.  
 CC -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC -!- SIMILARITY: Contains 1 SH2 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; L76687; AAC15861.1; -.  
 DR HSSP; P35235; 1AYA.  
 DR Genew; HGNC:4565; GRB14.  
 DR MIM; 601524; -.  
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA\_domain.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00788; RA; 1.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR ProDom; PD000093; SH2; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00314; RA; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50200; RA; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 KW SH2 domain; Phosphorylation.  
 FT DOMAIN 106 192 RAS-ASSOCIATING.  
 FT DOMAIN 234 342 PH.  
 FT DOMAIN 439 535 SH2.  
 SQ SEQUENCE 540 AA; 60954 MW; A8FCFC16D7437B47 CRC64;

Query Match 100.0%; Score 212; DB 1; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-20;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |||||  
 Db 367 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 409

# RESULT 2

GRBE\_MOUSE

ID GRBE\_MOUSE \* STANDARD; PRT; 538 AA.  
 AC Q9JLM9; Q8VDI2; Q9CR03;



DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Growth factor receptor-bound protein 14 (GRB14 adapter protein).  
 GN GRB14.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20179877; PubMed=10713090;  
 RA Reilly J.F., Mickey G., Maher P.A.;  
 RT "Association of fibroblast growth factor receptor 1 with the adaptor  
 RT protein Grb14. Characterization of a new receptor binding partner.";  
 RL J. Biol. Chem. 275:7771-7778(2000).  
 RN [2]  
 RP SEQUENCE OF 1-290 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic liver;  
 RX PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasaki Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [3]  
 RP SEQUENCE OF 332-538 FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary gland;  
 RX PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE  
 CC AUTOPHOSPHORYLATED INSULIN RECEPTOR WHICH IS THEN INHIBITED. THE  
 CC INTERACTION IS MEDIATED BY THE SH2 DOMAIN (By similarity).  
 CC -!- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-  
 CC terminus (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and  
 CC endosomes (By similarity).  
 CC -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES (BY SIMILARITY).  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC -!- SIMILARITY: Contains 1 SH2 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF155647; AAF43996.1; -.  
 DR EMBL; AK010849; BAB27221.2; -.  
 DR EMBL; AK010903; BAB27256.2; -.  
 DR EMBL; BC021820; AAH21820.1; -.  
 DR HSSP; P35235; 1AYA.  
 DR MGD; MGI:1355324; Grb14.  
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; IPI.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA\_domain.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00788; RA; 1.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR ProDom; PD000093; SH2; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00314; RA; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50200; RA; 1.

DR PROSITE; PS50001; SH2; 1.  
 KW SH2 domain; Phosphorylation.  
 FT DOMAIN 104 190 RAS-ASSOCIATING.  
 FT DOMAIN 232 340 PH.  
 FT DOMAIN 437 533 SH2.  
 SQ SEQUENCE 538 AA; 60573 MW; 04ABD6CEB6ABC6CB CRC64;

Query Match 96.7%; Score 205; DB 1; Length 538;  
 Best Local Similarity 93.0%; Pred. No. 2.1e-19;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||:|||||||:||||:|||||||  
 Db 365 PMRSVSENSLVAMDFSGEKSVIDNPTEALSVAVEEGLAWRKK 407

# RESULT 3

## GRBE\_RAT

ID GRBE\_RAT STANDARD; PRT; 538 AA.  
 AC O88900;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Growth factor receptor-bound protein 14 (GRB14 adapter protein).  
 GN GRB14.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar;  
 RX MEDLINE=98421528; PubMed=9748281;  
 RA Kasus-Jacobi A., Perdureau D., Auzan C., Clauser E., van Obberghen E.,  
 RA Mauvais-Jarvis F., Girard J., Burnol A.-F.;  
 RT "Identification of the rat adapter Grb14 as an inhibitor of insulin  
 RT actions.";  
 RL J. Biol. Chem. 273:26026-26035(1998).  
 CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE  
 CC AUTOPHOSPHORYLATED INSULIN RECEPTOR WHICH IS THEN INHIBITED. THE  
 CC INTERACTION IS MEDIATED BY THE SH2 DOMAIN.  
 CC -!- SUBUNIT: Binds to the ankyrin repeat region of TNKL via its N-  
 CC terminus (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and  
 CC endosomes (By similarity).  
 CC -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES (BY SIMILARITY).  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC -!- SIMILARITY: Contains 1 SH2 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF076619; AAC61478.1; -.

DR HSSP; P35235; 1AYA.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000159; RA\_domain.

DR InterPro; IPR000980; SH2.

DR Pfam; PF00169; PH; 1.

DR Pfam; PF00788; RA; 1.

DR Pfam; PF00017; SH2; 1.

DR PRINTS; PR00401; SH2DOMAIN.

DR ProDom; PD000093; SH2; 1.

DR SMART; SM00233; PH; 1.

DR SMART; SM00314; RA; 1.

DR SMART; SM00252; SH2; 1.

DR PROSITE; PS50003; PH\_DOMAIN; 1.

DR PROSITE; PS50200; RA; 1.

DR PROSITE; PS50001; SH2; 1.

KW SH2 domain; Phosphorylation.

FT DOMAIN 104 190 RAS-ASSOCIATING.

FT DOMAIN 232 340 PH.

FT DOMAIN 437 533 SH2.

SQ SEQUENCE 538 AA; 60592 MW; CEBC9037E7868EEF CRC64;

Query Match 96.7%; Score 205; DB 1; Length 538;  
 Best Local Similarity 93.0%; Pred. No. 2.1e-19;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||||:|||||||:||||:|||||||  
 Db 365 PMRSVSENSLVAMDFSGQKTRVIDNPTEALSVAVEEGLAWRKK 407

#### RESULT 4

##### GRBA\_HUMAN

ID GRBA\_HUMAN STANDARD; PRT; 594 AA.

AC Q13322; O00427; O00701; O75222; Q92606; Q92907; Q92948;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Growth factor receptor-bound protein 10 (GRB10 adaptor protein)

DE (Insulin receptor binding protein GRB-IR).

GN GRB10 OR GRBIR OR KIAA0207.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Skeletal muscle;

RX MEDLINE=96036069; PubMed=7479769;

RA Liu F., Roth R.A.;

RT "Grb-IR: a SH2-domain-containing protein that binds to the insulin receptor and inhibits its function.";

RL Proc. Natl. Acad. Sci. U.S.A. 92:10287-10291(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;  
 RA Nantel A., Mohammad-Ali K., Sherk J., Posner B.I., Thomas D.Y.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
 RX MEDLINE=99096036; PubMed=9881709;  
 RA Angrist M., Bolk S., Bentley K., Nallasamy S., Halushka M.K.,  
 RA Chakravarti A.;  
 RT "Genomic structure of the gene for the SH2 and pleckstrin homology  
 RT domain-containing protein GRB10 and evaluation of its role in  
 RT Hirschsprung disease.";  
 RL Oncogene 17:3065-3070(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,  
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI.  
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and brain.";  
 RL DNA Res. 3:321-329(1996).  
 RN [5]  
 RP SEQUENCE OF 1-398 FROM N.A.  
 RA Dauphin S., Biewald T.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Cerebellum, and Skeletal muscle;  
 RX MEDLINE=97160567; PubMed=9006901;  
 RA Frantz J.D., Giorgetti-Peraldi S., Ottinger E.A., Shoelson S.E.;  
 RT "Human GRB-IR-beta/GRB10: splice variants of an insulin and growth  
 RT factor receptor-binding protein with PH and SH2 domains.";  
 RL J. Biol. Chem. 272:2659-2667(1997).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=96394311; PubMed=8798417;  
 RA O'Neill T.J., Rose D.W., Pillay T.S., Hotta K., Olefsky J.M.,  
 RA Gustafson T.A.;  
 RT "Interaction of a GRB-IR splice variant (a human GRB10 homolog) with  
 RT the insulin and insulin-like growth factor I receptors. Evidence for  
 RT a role in mitogenic signaling.";  
 RL J. Biol. Chem. 271:22506-22513(1996).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=20320688; PubMed=10861285;  
 RA Blagitko N., Mergenthaler S., Schulz U., Wollmann H.A., Craigen W.,  
 RA Eggermann T., Ropers H.-H., Kalscheuer V.M.;  
 RT "Human GRB10 is imprinted and expressed from the paternal and maternal  
 RT allele in a highly tissue- and isoform-specific fashion.";  
 RL Hum. Mol. Genet. 9:1587-1595(2000).  
 CC -!- FUNCTION: PLAYS A FUNCTIONAL ROLE IN INSULIN AND IGF-I SIGNALING.  
 CC MAY SERVE TO POSITIVELY LINK THE INSULIN AND IGF-I RECEPTORS TO AN  
 CC UNCHARACTERIZED MITOGENIC SIGNALING PATHWAY. INTERACTS WITH THE  
 CC CYTOPLASMIC DOMAIN OF THE AUTOPHOSPHORYLATED INSULIN RECEPTOR

CC WHICH IS THEN INHIBITED. THE INTERACTION IS MEDIATED BY THE SH2  
 CC DOMAIN. ALSO BINDS ACTIVATED PLATELET-DERIVED GROWTH FACTOR  
 CC RECEPTOR AND EPIDERMAL GROWTH FACTOR RECEPTOR.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=3; Synonyms=Zeta;  
 CC IsoId=Q13322-1; Sequence=Displayed;  
 CC Name=1; Synonyms=Alpha;  
 CC IsoId=Q13322-2; Sequence=VSP\_001843;  
 CC Name=2; Synonyms=Beta, SV-1;  
 CC IsoId=Q13322-3; Sequence=VSP\_001842;  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SKELETAL MUSCLE.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC -!- SIMILARITY: Contains 1 SH2 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U34355; AAA88819.1; -.  
 DR EMBL; AF000017; AAC19748.1; -.  
 DR EMBL; AF073378; AAC83655.1; -.  
 DR EMBL; AF073363; AAC83655.1; JOINED.  
 DR EMBL; AF073364; AAC83655.1; JOINED.  
 DR EMBL; AF073365; AAC83655.1; JOINED.  
 DR EMBL; AF073366; AAC83655.1; JOINED.  
 DR EMBL; AF073367; AAC83655.1; JOINED.  
 DR EMBL; AF073368; AAC83655.1; JOINED.  
 DR EMBL; AF073369; AAC83655.1; JOINED.  
 DR EMBL; AF073370; AAC83655.1; JOINED.  
 DR EMBL; AF073371; AAC83655.1; JOINED.  
 DR EMBL; AF073372; AAC83655.1; JOINED.  
 DR EMBL; AF073373; AAC83655.1; JOINED.  
 DR EMBL; AF073374; AAC83655.1; JOINED.  
 DR EMBL; AF073375; AAC83655.1; JOINED.  
 DR EMBL; AF073376; AAC83655.1; JOINED.  
 DR EMBL; AF073377; AAC83655.1; JOINED.  
 DR EMBL; AF073378; AAC83654.1; -.  
 DR EMBL; AF073363; AAC83654.1; JOINED.  
 DR EMBL; AF073364; AAC83654.1; JOINED.  
 DR EMBL; AF073365; AAC83654.1; JOINED.  
 DR EMBL; AF073366; AAC83654.1; JOINED.  
 DR EMBL; AF073367; AAC83654.1; JOINED.  
 DR EMBL; AF073368; AAC83654.1; JOINED.  
 DR EMBL; AF073369; AAC83654.1; JOINED.  
 DR EMBL; AF073371; AAC83654.1; JOINED.  
 DR EMBL; AF073372; AAC83654.1; JOINED.  
 DR EMBL; AF073373; AAC83654.1; JOINED.  
 DR EMBL; AF073374; AAC83654.1; JOINED.  
 DR EMBL; AF073375; AAC83654.1; JOINED.

DR EMBL; AF073376; AAC83654.1; JOINED.  
 DR EMBL; AF073377; AAC83654.1; JOINED.  
 DR EMBL; D86962; BAA13198.1; -.  
 DR EMBL; AF001534; AAB81134.1; -.  
 DR EMBL; AC005153; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; U69276; AAB08431.1; -.  
 DR EMBL; U66065; AAC50671.1; -.  
 DR EMBL; AJ271366; CAB96542.1; -.  
 DR PIR; I39175; I39175.  
 DR HSSP; O60880; 1D1Z.  
 DR Genew; HGNC:4564; GRB10.  
 DR MIM; 601523; -.  
 DR GO; GO:0005737; C:cytoplasm; TAS.  
 DR GO; GO:0005886; C:plasma membrane; TAS.  
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.  
 DR GO; GO:0007267; P:cell-cell signaling; TAS.  
 DR GO; GO:0008286; P:insulin receptor signaling pathway; TAS.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA domain.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00788; RA; 1.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR ProDom; PD000093; SH2; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00314; RA; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50200; RA; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 KW SH2 domain; Alternative splicing.  
 FT DOMAIN 166 250 RAS-ASSOCIATING.  
 FT DOMAIN 290 399 PH.  
 FT DOMAIN 493 574 SH2.  
 FT VARSPLIC 1 58 Missing (in isoform 2).  
 FT /FTId=VSP\_001842.  
 FT VARSPLIC 283 328 Missing (in isoform 1).  
 FT /FTId=VSP\_001843.  
 FT CONFLICT 1 17 MALAGCPDSFLHHPYYQ -> MQAAGPLFRSK (IN REF.  
 FT 4).  
 FT CONFLICT 152 152 P -> A (IN REF. 4).  
 FT CONFLICT 400 400 G -> E (IN REF. 6).  
 FT CONFLICT 498 498 I -> F (IN REF. 6).  
 FT CONFLICT 541 541 N -> I (IN REF. 7).  
 SQ SEQUENCE 594 AA; 67231 MW; 53A5F885E17C6C6B CRC64;

Query Match 79.7%; Score 169; DB 1; Length 594;  
 Best Local Similarity 76.7%; Pred. No. 1.3e-14;  
 Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |:||:||||||| ||||| || |:||| |||:  
 Db 423 PVRSVSENSLVAMDFSGQTGRVIENPAEAQSAALEEGHAWRKR 465

RESULT 5

## GRB7\_HUMAN

ID GRB7\_HUMAN STANDARD; PRT; 532 AA.  
AC Q14451; Q92568; Q96DF9;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Growth factor receptor-bound protein 7 (GRB7 adapter protein)  
DE (Epidermal growth factor receptor GRB-7) (B47).  
GN GRB7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oesophageal carcinoma;  
RX MEDLINE=97236270; PubMed=9125150;  
RA Kishi T., Sasaki H., Akiyama N., Ishizuka T., Sakamoto H., Aizawa S.,  
RA Sugimura T., Terada M.;  
RT "Molecular cloning of human GRB-7 co-amplified with CAB1 and c-ERBB-2  
RT in primary gastric cancer.";  
RL Biochem. Biophys. Res. Commun. 232:5-9(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98376491; PubMed=9710451;  
RA Tanaka S., Mori M., Akiyoshi T., Tanaka Y., Mafune K., Wands J.R.,  
RA Sugimachi K.;  
RT "A novel variant of human Grb7 is associated with invasive esophageal  
RT carcinoma.";  
RL J. Clin. Invest. 102:821-827(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Whittock N.V., Eady R.A.J., McGrath J.A.;  
RT "Genomic organization and amplification of the human GRB7 gene.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length



RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE OF 130-343 FROM N.A.  
 RX MEDLINE=97141776; PubMed=8988034;  
 RA Tanaka S., Mori M., Akiyoshi T., Tanaka Y., Mafune K., Wands J.R.,  
 RA Sugimachi K.;  
 RT "Coexpression of Grb7 with epidermal growth factor receptor or  
 RT Her2/erbB2 in human advanced esophageal carcinoma.";  
 RL Cancer Res. 57:28-31(1997).  
 CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE EPIDERMAL  
 CC GROWTH FACTOR RECEPTOR WHICH IS THEN INHIBITED. THE INTERACTION IS  
 CC MEDIATED BY THE SH2 DOMAIN. ALSO BINDS TO ERBB2.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=1;  
 CC Comment=At least 2 isoforms are produced;  
 CC Name=1;  
 CC IsoId=Q14451-1; Sequence=Displayed;  
 CC -!- PTM: PHOSPHORYLATED ON TYROSINE RESIDUES.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC -!- SIMILARITY: Contains 1 SH2 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D43772; BAA07827.1; -.  
 DR EMBL; AB008789; BAA29059.1; -.  
 DR EMBL; AF274875; AAG25938.1; -.  
 DR EMBL; BC006535; AAH06535.1; -.  
 DR EMBL; D87513; BAA13412.1; -.  
 DR PIR; JC5412; JC5412.  
 DR HSSP; P00519; 1AB2.  
 DR Genew; HGNC:4567; GRB7.  
 DR MIM; 601522; -.  
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.  
 DR GO; GO:0007173; P:EGF receptor signaling pathway; TAS.  
 DR GO; GO:0007048; P:oncogenesis; TAS.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA\_domain.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00788; RA; 1.  
 DR Pfam; PF00017; SH2; 1.  
 DR ProDom; PD000093; SH2; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00314; RA; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50200; RA; 1.  
 DR PROSITE; PS50001; SH2; 1.

KW	SH2 domain; Phosphorylation; Alternative splicing.			
FT	DOMAIN	100	186	RAS-ASSOCIATING.
FT	DOMAIN	229	338	PH.
FT	DOMAIN	431	512	SH2.
FT	CONFLICT	18	18	W -> C (IN REF. 4).
SO	SEQUENCE 532 AA; 59764 MW; A68679F83A146F74 CRC64:			

Query Match 76.4%; Score 162; DB 1; Length 532;  
Best Local Similarity 74.4%; Pred. No. 9.9e-14;  
Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy            1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
| : | | : : | | | | | | | | | | | | | | | | | | | |  
Db         363 PLRSASDNTLTVAMDFSGHAGRVIENPREALSVALEEAQAWRKK 405

RESULT 6

GRBA MOUSE

```

ID      GRBA MOUSE          STANDARD;          PRT;      621 AA.
AC      Q60760; O35352;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Growth factor receptor-bound protein 10 (GRB10 adaptor protein).
GN      GRB10.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Swiss;
RX      MEDLINE=95249278; PubMed=7731717;
RA      Ooi J., Yajnik V., Immanuel D., Gordon M., Moskow J.J., Buchberg A.,
RA      Margolis B.;
RT      "The cloning of Grb10 reveals a new family of SH2 domain proteins.";
RL      Oncogene 10:1621-1630(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97216049; PubMed=9062339;
RA      Laviola L., Giorgino F., Chow J.C., Baquero J.A., Hansen H., Ooi J.,
RA      Zhu J., Riedel H., Smith R.J.;
RT      "The adapter protein Grb10 associates preferentially with the insulin
RT      receptor as compared with the IGF-I receptor in mouse fibroblasts.";
RL      J. Clin. Invest. 99:830-837(1997).
CC      -!- FUNCTION: PLAYS A FUNCTIONAL ROLE IN INSULIN AND IGF-I SIGNALING.
CC      MAY SERVE TO POSITIVELY LINK THE INSULIN AND IGF-I RECEPTORS TO AN
CC      UNCHARACTERIZED MITOGENIC SIGNALING PATHWAY. INTERACTS WITH THE
CC      CYTOPLASMIC DOMAIN OF THE AUTOPHOSPHORYLATED INSULIN RECEPTOR
CC      WHICH IS THEN INHIBITED. THE INTERACTION IS MEDIATED BY THE SH2
CC      DOMAIN. ALSO BINDS ACTIVATED PLATELET-DERIVED GROWTH FACTOR
CC      RECEPTOR AND EPIDERMAL GROWTH FACTOR RECEPTOR.
CC      -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=2;
CC      Name=1;
CC      IsoId=Q60760-1; Sequence=Displayed;
CC      Name=2;

```

CC           IsoId=Q60760-2; Sequence=VSP\_001844;  
 CC    -!- SIMILARITY: Contains 1 PH domain.  
 CC    -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC    -!- SIMILARITY: Contains 1 SH2 domain.  
 CC    -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.  
 CC    -----  
 CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC    the European Bioinformatics Institute. There are no restrictions on its  
 CC    use by non-profit institutions as long as its content is in no way  
 CC    modified and this statement is not removed. Usage by and for commercial  
 CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC    -----  
 DR    EMBL; U18996; AAB53687.1; -.  
 DR    EMBL; AF022072; AAB72103.1; -.  
 DR    PIR; I49199; I49199.  
 DR    HSSP; O60880; 1D1Z.  
 DR    MGD; MGI:103232; Grb10.  
 DR    GO; GO:0005070; F:SH3/SH2 adaptor protein activity; IPI.  
 DR    InterPro; IPR001849; PH.  
 DR    InterPro; IPR000159; RA\_domain.  
 DR    InterPro; IPR000980; SH2.  
 DR    Pfam; PF00169; PH; 1.  
 DR    Pfam; PF00788; RA; 1.  
 DR    Pfam; PF00017; SH2; 1.  
 DR    PRINTS; PR00401; SH2DOMAIN.  
 DR    ProDom; PD000093; SH2; 1.  
 DR    SMART; SM00233; PH; 1.  
 DR    SMART; SM00314; RA; 1.  
 DR    SMART; SM00252; SH2; 1.  
 DR    PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR    PROSITE; PS50200; RA; 1.  
 DR    PROSITE; PS50001; SH2; 1.  
 KW    SH2 domain; Alternative splicing.  
 FT    DOMAIN       194     278       RAS-ASSOCIATING.  
 FT    DOMAIN       318     427       PH.  
 FT    DOMAIN       520     601       SH2.  
 FT    VARSPPLIC   117     141       Missing (in isoform 2).  
 FT                               /FTid=VSP\_001844.  
 FT    CONFLICT     491     492       NG -> KR (IN REF. 2).  
 SQ    SEQUENCE    621 AA; 70471 MW; 2A9A45D5842468A7 CRC64;

Query Match                   75.9%; Score 161; DB 1; Length 621;  
 Best Local Similarity       78.0%; Pred. No. 1.6e-13;  
 Matches   32; Conservative   3; Mismatches   6; Indels    0; Gaps    0;

QY           1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
               ||||:|||||||       |||:|| || | |:||||  
 Db           450 PMRSVSENSLVAMDFSGQIGRVIDNPAAEQSAALEEGHAWR 490

# RESULT 7

GRB7\_MOUSE

ID   GRB7\_MOUSE       STANDARD;       PRT;   535 AA.

AC   Q03160;

DT   15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Growth factor receptor-bound protein 7 (GRB7 adapter protein)  
 DE (Epidermal growth factor receptor GRB-7).  
 GN GRB7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93028373; PubMed=1409582;  
 RA Margolis B., Silvennoinen O., Comoglio F., Roonprapunt C.,  
 RA Skolnik E.Y., Ullrich A., Schlessinger J.;  
 RT "High-efficiency expression/cloning of epidermal growth factor-  
 RT receptor-binding proteins with Src homology 2 domains.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8894-8898(1992).  
 CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC DOMAIN OF THE EPIDERMAL  
 CC GROWTH FACTOR RECEPTOR WHICH IS THEN INHIBITED. THE INTERACTION IS  
 CC MEDIATED BY THE SH2 DOMAIN. ALSO BINDS TO ERBB2.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
 CC -!- SIMILARITY: Contains 1 SH2 domain.  
 CC -!- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M94450; AAA37733.1; -.  
 DR PIR; C46243; C46243.  
 DR HSSP; P35235; 1AYA.  
 DR MGD; MGI:102683; Grb7.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000159; RA\_domain.  
 DR InterPro; IPR000980; SH2.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00017; SH2; 1.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR ProDom; PD000093; SH2; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00314; RA; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50200; RA; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 KW SH2 domain.  
 FT DOMAIN 99 185 RAS-ASSOCIATING.  
 FT DOMAIN 228 341 PH.  
 FT DOMAIN 434 515 SH2.  
 SQ SEQUENCE 535 AA; 59959 MW; CD8C307864703645 CRC64;

QY           1 PMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK     43  
             | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db          366 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKK   408

Qy           7 ENSLVAMDFSGQKSRV-IENPTEALSVAVEEGLAWRK 42  
              : | | : | : | | | : : || : | | |

## RESULT 9

## PAAY\_ECOLI

ID PAAY\_ECOLI STANDARD; PRT; 196 AA.

AC P77181; 053020;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Phenylacetic acid degradation protein paaY.

GN PAAY OR B1400.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=W / ATCC 11105;

RX MEDLINE=98421522; PubMed=9748275;

RA Ferrandez A., Minambres B., Garcia B., Olivera E.R., Luengo J.M.,

RA Garcia J.L., Diaz E.;

RT "Catabolism of phenylacetic acid in Escherichia coli. Characterization  
RT of a new aerobic hybrid pathway.";

RL J. Biol. Chem. 273:25974-25986(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97251357; PubMed=9097039;

RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,

RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,

RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,

RA Oshima T., Saito N., Sampei G., Seki Y., Sivasundaram S.,

RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,

RA Yamamoto Y., Horiuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

CC -!- PATHWAY: Phenylacetic acid aerobic catabolism.

CC -!- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF

CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).

CC

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X97452; CAA66102.1; -.  
DR EMBL; AE000237; AAC74482.1; -.  
DR EMBL; D90778; BAA15008.1; -.  
DR PIR; C64891; C64891.  
DR EcoGene; EG13747; paaY.  
DR InterPro; IPR001451; Hexapep\_transf.  
DR Pfam; PF00132; hexapep; 4.  
DR PROSITE; PS00101; HEXAPEP\_TRANSFERASES; FALSE\_NEG.  
KW Transferase; Repeat; Complete proteome.  
FT VARIANT 75 75 G -> E (IN STRAIN W).  
FT VARIANT 179 179 I -> V (IN STRAIN W).  
FT VARIANT 182 182 G -> N (IN STRAIN W).  
SQ SEQUENCE 196 AA; 21324 MW; FA3454F5AA0910DB CRC64;

Query Match 25.7%; Score 54.5; DB 1; Length 196;  
Best Local Similarity 32.6%; Pred. No. 5.4;  
Matches 15; Conservative 10; Mismatches 14; Indels 7; Gaps 2;

Qy 5 ISENSLV-AMDFSGQKSR-----VIENPTEALSVAVEEGLAWRKK 43  
| |||:| | | : :: :| :|: | : |||:|:  
Db 109 IGENSIVGASAFVKAKAEMPANYLIVGSPAKAIRELSEQELAWKKQ 154

#### RESULT 10

##### HIS6\_METTH

ID HIS6\_METTH STANDARD; PRT; 274 AA.  
AC O27398;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP  
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase  
DE subunit hisF) (IGPS subunit hisF).  
GN HISF OR MTH1343.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics."  
RL J. Bacteriol. 179:7135-7155(1997).  
CC -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to

CC IGP, AICAR and glutamate. The hisF subunit catalyzes the  
 CC cyclization activity that produces IGP and AICAR from PRFAR using  
 CC the ammonia provided by the hisH subunit (By similarity).  
 CC -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-  
 CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-  
 CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-  
 CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.  
 CC -!- PATHWAY: Histidine biosynthesis; fifth step.  
 CC -!- SUBUNIT: Heterodimer of hisH and hisF (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000897; AAB85821.1; ALT\_INIT.  
 DR HAMAP; MF\_01013; -, 1.  
 DR InterPro; IPR003009; FMN\_enzyme.  
 DR InterPro; IPR006062; His\_biosynth.  
 DR InterPro; IPR004651; HisF.  
 DR Pfam; PF00977; His\_biosynth; 1.  
 DR TIGRFAMs; TIGR00735; hisF; 1.  
 KW Histidine biosynthesis; Lyase; Complete proteome.  
 FT ACT\_SITE 11 11 POTENTIAL.  
 FT ACT\_SITE 134 134 POTENTIAL.  
 SQ SEQUENCE 274 AA; 30463 MW; B80082BE4552AC53 CRC64;

Query Match 25.5%; Score 54; DB 1; Length 274;  
 Best Local Similarity 36.8%; Pred. No. 9;  
 Matches 14; Conservative 7; Mismatches 11; Indels 6; Gaps 2;

QY 6 SENSLVAMDFSGQKSRVIENPTEA---LSVAVEEGLAW 40  
 |: :||:| | | ||| |: :|::| |  
 Db 126 SQACVVAID---AKRRYIENPRESDERFIEVDDGYCW 160

# RESULT 11

## ENO\_SULTO

ID ENO\_SULTO STANDARD; PRT; 416 AA.  
 AC Q972B6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-  
 DE glycerate hydro-lyase).  
 GN ENO OR ST1212.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RC STRAIN=JCM 10545 / 7;  
 RX MEDLINE=21456156; PubMed=11572479;  
 RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermoacidophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 CC -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate +  
 CC H(2)O.  
 CC -!- COFACTOR: Magnesium is required for catalysis and for stabilizing  
 CC the dimer (By similarity).  
 CC -!- PATHWAY: Glycolysis.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP000985; BAB66253.1; -.  
 DR HAMAP; MF\_00318; -; 1.  
 DR InterPro; IPR000941; Enolase.  
 DR Pfam; PF00113; enolase; 1.  
 DR Pfam; PF03952; enolase\_N; 1.  
 DR PRINTS; PR00148; ENOLASE.  
 DR ProDom; PD000902; Enolase; 1.  
 DR TIGRFAMs; TIGR01060; eno; 1.  
 DR PROSITE; PS00164; ENOLASE; FALSE\_NEG.  
 KW Lyase; Glycolysis; Magnesium; Complete proteome.  
 FT ACT\_SITE 152 152 BY SIMILARITY.  
 FT METAL 239 239 MAGNESIUM (BY SIMILARITY).  
 FT METAL 280 280 MAGNESIUM (BY SIMILARITY).  
 FT METAL 306 306 MAGNESIUM (BY SIMILARITY).  
 SQ SEQUENCE 416 AA; 46304 MW; 3E480E37CD434815 CRC64;

Query Match 25.5%; Score 54; DB 1; Length 416;  
 Best Local Similarity 42.4%; Pred. No. 14;  
 Matches 14; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 7 ENSLVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
 : : | : | | | | | : | : | : |  
 Db 88 DQTLIRMDGTPNKS RVGGNTTIATSI AVAKTAA 120

RESULT 12  
 ENO\_AERPE  
 ID ENO\_AERPE STANDARD; PRT; 432 AA.  
 AC Q9Y927;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase).  
 GN ENO OR APE2458.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
 OC Desulfurococcaceae; Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1.";  
 RL DNA Res. 6:83-101(1999).  
 CC -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate +  
 CC H(2)O.  
 CC -!- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING  
 CC THE DIMER (BY SIMILARITY).  
 CC -!- PATHWAY: Glycolysis.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP000064; BAA81473.1; -.  
 DR PIR; A72477; A72477.  
 DR HSSP; P00924; 4ENL.  
 DR HAMAP; MF\_00318; -; 1.  
 DR InterPro; IPR000941; Enolase.  
 DR Pfam; PF00113; enolase; 1.  
 DR Pfam; PF03952; enolase\_N; 1.  
 DR PRINTS; PR00148; ENOLASE.  
 DR ProDom; PD000902; Enolase; 1.  
 DR TIGRFAMs; TIGR01060; eno; 1.  
 DR PROSITE; PS00164; ENOLASE; FALSE\_NEG.  
 KW Lyase; Glycolysis; Magnesium; Complete proteome.  
 FT ACT\_SITE 158 158 BY SIMILARITY.  
 FT METAL 247 247 MAGNESIUM (BY SIMILARITY).  
 FT METAL 288 288 MAGNESIUM (BY SIMILARITY).  
 FT METAL 315 315 MAGNESIUM (BY SIMILARITY).  
 SQ SEQUENCE 432 AA; 46344 MW; 924E6362F8BDFDDE CRC64;

Query Match 25.5%; Score 54; DB 1; Length 432;  
 Best Local Similarity 43.3%; Pred. No. 15;  
 Matches 13; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 10 LVAMDFSGQKSRVIENPTEALSVAVEEGLA 39  
 | : : | : ||| : | | ||| : || |  
 Db 97 LIELDGTTPNKSRLGGNTTTALSIAVSRAAA 126

# RESULT 13

GUAA\_BACHD

ID GUAA\_BACHD STANDARD; PRT; 513 AA.

AC Q9KF78;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine  
 DE amidotransferase) (GMP synthetase).

GN GUAA OR BH0607.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.";

RL Nucleic Acids Res. 28:4317-4331(2000).

CC -!- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +  
 CC H(2)O = AMP + diphosphate + GMP + L-glutamate.

CC -!- PATHWAY: GMP biosynthesis.

CC -!- SUBUNIT: Homodimer (By similarity).

CC -!- MISCELLANEOUS: THE HISTIDINE EXPECTED IN POSITION 172 AND REQUIRED  
 CC FOR THE ACTIVE SITE IS MISSING.

CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE  
 CC FAMILY.

CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC -!- CAUTION: Could lack activity as the potential active site His  
 CC residue in position 172 is replaced by a Gln.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AP001509; BAB04326.1; -.

DR PIR; G83725; G83725.

DR HSSP; P04079; 1GPM.

DR HAMAP; MF\_00344; atypical; 1.

DR InterPro; IPR006220; Anth\_synthII.  
 DR InterPro; IPR001317; CPS\_GATase.  
 DR InterPro; IPR000991; GATase\_1.  
 DR InterPro; IPR001674; GMP\_synt\_C.  
 DR InterPro; IPR004739; GMPSynthase\_N.  
 DR Pfam; PF00117; GATase; 1.  
 DR Pfam; PF00958; GMP\_synt\_C; 1.  
 DR PRINTS; PR00097; ANTSNTHASEII.  
 DR PRINTS; PR00099; CPSGATASE.  
 DR PRINTS; PR00096; GATASE.  
 DR TIGRFAMs; TIGR00884; guaA\_Cterm; 1.  
 DR TIGRFAMs; TIGR00888; guaA\_Nterm; 1.  
 DR PROSITE; PS00442; GATASE\_TYPE\_I; 1.  
 KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;  
 KW Glutamine amidotransferase; Complete proteome.  
 FT DOMAIN 1 197 GLUTAMINE AMIDOTRANSFERASE.  
 FT DOMAIN 230 389 GMP-BINDING (BY SIMILARITY).  
 FT ACT\_SITE 85 85 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 174 174 GATASE (BY SIMILARITY).  
 FT NP\_BIND 226 232 ATP (BY SIMILARITY).  
 SQ SEQUENCE 513 AA; 57328 MW; 1308CA1ED1923379 CRC64;

Query Match 25.5%; Score 54; DB 1; Length 513;  
 Best Local Similarity 35.3%; Pred. No. 18;  
 Matches 12; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 2 MRSISENSLVAMDFSGQKSRVIENPTEALSVAVE 35  
 | :|| :| :|| || ::| | | |  
 Db 1 MEQLSEEMIVVLDFGGQYNQLITRRIRDLGVYSE 34

#### RESULT 14

##### DLD1\_KLULA

ID DLD1\_KLULA STANDARD; PRT; 579 AA.  
 AC Q12627;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE D-lactate dehydrogenase [cytochrome], mitochondrial precursor  
 DE (EC 1.1.2.4) (D-lactate ferricytochrome C oxidoreductase) (D-LCR).  
 GN DLD1 OR DLD.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;  
 RX MEDLINE=95058916; PubMed=7969031;  
 RA Lodi T., O'Connor D., Goffrini P., Ferrero I.;  
 RT "Carbon catabolite repression in Kluyveromyces lactis: isolation and  
 RT characterization of the KIDL gene encoding the mitochondrial enzyme  
 RT D-lactate ferricytochrome c oxidoreductase.";  
 RL Mol. Gen. Genet. 244:622-629(1994).  
 CC -!- FUNCTION: CATALYZE THE STEREOSPECIFIC OXIDATION OF D-LACTATE TO  
 CC PYRUVATE.  
 CC -!- CATALYTIC ACTIVITY: (R)-lactate + 2 ferricytochrome c = pyruvate +

CC 2 ferrocytochrome c.  
 CC -!- COFACTOR: CONTAINS TWO FAD AND FOUR TO SIX ZINC MOLES PER MOLE.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -!- SIMILARITY: BELONGS TO THE FAD-BINDING OXIDOREDUCTASE/TRANSFERASE  
 CC FAMILY 4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X71628; CAA50635.1; -.  
 DR PIR; S51528; S51528.  
 DR InterPro; IPR004113; FAD-oxidase\_C.  
 DR InterPro; IPR006094; Oxid\_FAD\_bind.  
 DR Pfam; PF02913; FAD-oxidase\_C; 1.  
 DR Pfam; PF01565; FAD\_binding\_4; 1.  
 KW Oxidoreductase; Flavoprotein; FAD; Transit peptide; Mitochondrion;  
 KW Zinc.  
 FT TRANSIT 1 ? MITOCHONDRION.  
 FT CHAIN ? 579 D-LACTATE DEHYDROGENASE [CYTOCHROME].  
 SQ SEQUENCE 579 AA; 63484 MW; 0DE3A07DC4934883 CRC64;

Query Match 25.0%; Score 53; DB 1; Length 579;  
 Best Local Similarity 32.4%; Pred. No. 28;  
 Matches 11; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

Qy 9 SLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRK 42  
 | | : | | : : | : | | | : | : | :  
 Db 190 SCVVLDISKYLNKIIQLNKEDLDVVVQGGVPWEE 223

# RESULT 15

## NADA\_HELPY

ID NADA\_HELPY STANDARD; PRT; 336 AA.  
 AC O25910;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Quinolinate synthetase A.  
 GN NADA OR HP1356.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,  
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen *Helicobacter*  
RT *pylori*.";

RL Nature 388:539-547(1997).

CC -!- FUNCTION: Catalyzes the condensation of iminoaspartate with  
CC dihydroxyacetone phosphate to form quinolinate.

CC -!- PATHWAY: NAD biosynthesis; aspartate to NaMN; second step.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: BELONGS TO THE QUINOLINATE SYNTHETASE A FAMILY.  
CC SUBFAMILY 3.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE000636; AAD08398.1; -.

DR PIR; D64689; D64689.

DR TIGR; HP1356; -.

DR HAMAP; MF\_00569; -, 1.

DR InterPro; IPR003473; NadA.

DR Pfam; PF02445; NadA; 1.

DR TIGRFAMs; TIGR00550; nadA; 1.

KW Pyridine nucleotide biosynthesis; Complete proteome.

SQ SEQUENCE 336 AA; 37812 MW; 963569A848239C4F CRC64;

Query Match 24.5%; Score 52; DB 1; Length 336;

Best Local Similarity 34.9%; Pred. No. 21;

Matches 15; Conservative 9; Mismatches 13; Indels 6; Gaps 2;

Qy 7 ENSLVA-MDFSGQKSRVIE-----NPTEALSVAVEEGLAWRKK 43

| |:|: |||| |::|| :| : ::| | | |

Db 228 EPSVVSNAADFSGSTSQIIEFVEKLSPNQKVAIGTESHLVNRLK 270

#### RESULT 16

YPEB\_OCEIH

ID YPEB\_OCEIH STANDARD; PRT; 447 AA.

AC P59106;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sporulation protein ypeB.

GN OB1805.

OS *Oceanobacillus iheyensis*.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Oceanobacillus*.

OX NCBI\_TaxID=182710;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HTE831 / DSM 14371 / JCM 11309;

RX MEDLINE=22220767; PubMed=12235376;

RA Takami H., Takaki Y., Uchiyama I.;  
 RT "Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.";  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 CC -!- FUNCTION: Required for spore cortex hydrolysis during germination.  
 CC Appears to be required for either expression, localization,  
 CC activation or function of sleB (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE YPEB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP004599; BAC13761.1; -.  
 KW Sporulation; Germination; Complete proteome.  
 SQ SEQUENCE 447 AA; 50695 MW; FE260E8ED5932A5E CRC64;

Query Match 24.5%; Score 52; DB 1; Length 447;  
 Best Local Similarity 26.7%; Pred. No. 29;  
 Matches 12; Conservative 10; Mismatches 17; Indels 6; Gaps 1;

Qy 2 MRSISENSLV-----AMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
 :|:::| | | : | :: : : ||:|||| |  
 Db 116 VRNLDDNPLTEEETQKLKDYDQSGQIKDELQVQHVAALEEGLNW 160

RESULT 17  
 6PGD\_LACLC  
 ID 6PGD\_LACLC STANDARD; PRT; 472 AA.  
 AC P96789;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 6-phosphogluconate dehydrogenase (EC 1.1.1.44).  
 GN GND.  
 OS *Lactococcus lactis* (subsp. *cremoris*) (*Streptococcus cremoris*).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; *Lactococcus*.  
 OX NCBI\_TaxID=1359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MG1363;  
 RX MEDLINE=99131986; PubMed=9931298;  
 RA Tetaud E., Hanau S., Wells J.M., Le Page R.W.F., Adams M.J.,  
 RA Arkison S., Barrett M.P.;  
 RT "6-Phosphogluconate dehydrogenase from *Lactococcus lactis*: a role for  
 RT arginine residues in binding substrate and coenzyme.";  
 RL Biochem. J. 338:55-60(1999).  
 CC -!- CATALYTIC ACTIVITY: 6-phospho-D-gluconate + NADP(+) = D-ribulose  
 CC 5-phosphate + CO(2) + NADPH.  
 CC -!- PATHWAY: Hexose monophosphate shunt.  
 CC -!- SIMILARITY: BELONGS TO THE 6-PHOSPHOGLUCONATE DEHYDROGENASE  
 CC FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U74322; AAC12804.1; -.
DR HSSP; P00349; 2PGD.
DR InterPro; IPR006183; 6PGD.
DR InterPro; IPR006114; 6PGD_C.
DR InterPro; IPR006113; 6PGD_decarbox.
DR InterPro; IPR006115; 6PGD_NAD.
DR InterPro; IPR006184; 6PGdom.
DR Pfam; PF00393; 6PGD; 1.
DR Pfam; PF03446; NAD_binding_2; 1.
DR PRINTS; PR00076; 6PGDHDRGNASE.
DR TIGRFAMs; TIGR00873; gnd; 1.
DR PROSITE; PS00461; 6PGD; 1.
KW Gluconate utilization; Oxidoreductase; Pentose shunt; NADP.
SQ SEQUENCE 472 AA; 52444 MW; 739958A068D63CD0 CRC64;

```

```

Query Match          24.5%; Score 52; DB 1; Length 472;
Best Local Similarity 38.9%; Pred. No. 30;
Matches 14; Conservative 6; Mismatches 12; Indels 4; Gaps 1;

```

```

Qy      12 AMDFSGQKSRVIENPTEAL----SVAVEEGLAWRKK 43
      |:|||| | ||| :|| :: :| | :|
Db      309 ALDFSGDKKEVIEKIRKALYFSKIMSYAQGFAQLRK 344

```

# RESULT 18

## NUSG\_TREPA

```

ID NUSG_TREPA STANDARD; PRT; 185 AA.
AC O83264;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription antitermination protein nusG.
GN NUSG OR TP0236.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis

```



RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC -!- FUNCTION: INFLUENCES TRANSCRIPTION TERMINATION AND  
 CC ANTITERMINATION. ACTS AS A COMPONENT OF THE TRANSCRIPTION COMPLEX,  
 CC AND INTERACTS WITH THE TERMINATION FACTOR RHO AND RNA POLYMERASE  
 CC (BY SIMILARITY).  
 CC -!- SIMILARITY: Belongs to the nusG family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001205; AAC65224.1; -.  
 DR PIR; F71349; F71349.  
 DR TIGR; TP0236; -.  
 DR InterPro; IPR005824; KOW.  
 DR InterPro; IPR006646; KOW\_sub.  
 DR InterPro; IPR006645; NgN.  
 DR InterPro; IPR001062; NusG.  
 DR Pfam; PF00467; KOW; 1.  
 DR Pfam; PF02357; NusG; 1.  
 DR PRINTS; PR00338; NUSGTNSCPFCT.  
 DR SMART; SM00739; KOW; 1.  
 DR SMART; SM00738; NGN; 1.  
 DR TIGRFAMs; TIGR00922; nusG; 1.  
 DR PROSITE; PS01014; NUSG; 1.  
 KW Transcription termination; Complete proteome.  
 SQ SEQUENCE 185 AA; 20928 MW; DF9DB89A4A2F9F52 CRC64;

Query Match 24.3%; Score 51.5; DB 1; Length 185;  
 Best Local Similarity 35.9%; Pred. No. 13;  
 Matches 14; Conservative 7; Mismatches 13; Indels 5; Gaps 1;

QY 5 ISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 |:: || ||: |::| | ||| :: ||  
 Db 128 IAQTFLV-----GQQVRIVEGPFATFSGEVEEVMSEKRNK 161

# RESULT 19

CARA\_THEME  
 ID CARA\_THEME STANDARD; PRT; 392 AA.  
 AC Q9WZ28;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Carbamoyl-phosphate synthase small chain (EC 6.3.5.5) (Carbamoyl-  
 DE phosphate synthetase glutamine chain).  
 GN CARA OR TM0558.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of *Thermotoga maritima*."  
 RL Nature 399:323-329(1999).  
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +  
 CC phosphate + L-glutamate + carbamoyl phosphate.  
 CC -!- PATHWAY: Arginine biosynthesis.  
 CC -!- PATHWAY: Pyrimidine biosynthesis; first step.  
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain  
 CC promotes the hydrolysis of glutamine to ammonia, which is used by  
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the carA family.  
 CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001730; AAD35643.1; -.  
 DR PIR; D72363; D72363.  
 DR HSSP; P00907; 1CS0.  
 DR TIGR; TM0558; -.  
 DR HAMAP; MF\_01209; atypical; 1.  
 DR InterPro; IPR006274; CarA\_small.  
 DR InterPro; IPR001317; CPS\_GATase.  
 DR InterPro; IPR002474; CPSase\_sm\_chain.  
 DR InterPro; IPR000991; GATase\_1.  
 DR Pfam; PF00988; CPSase\_sm\_chain; 1.  
 DR Pfam; PF00117; GATase; 1.  
 DR PRINTS; PR00099; CPSGATASE.  
 DR PRINTS; PR00096; GATASE.  
 DR TIGRFAMs; TIGR01368; CPSaseI\_small; 1.  
 DR PROSITE; PS00442; GATASE\_TYPE\_I; FALSE\_NEG.  
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase;  
 KW Glutamine amidotransferase; Complete proteome.  
 FT DOMAIN 1 176 CPSASE.  
 FT DOMAIN 177 392 GLUTAMINE AMIDOTRANSFERASE.  
 FT ACT\_SITE 252 252 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 392 AA; 42930 MW; B5312FBB07B181FC CRC64;

Query Match 24.3%; Score 51.5; DB 1; Length 392;  
 Best Local Similarity 31.8%; Pred. No. 29;  
 Matches 14; Conservative 10; Mismatches 15; Indels 5; Gaps 3;

QY 2 MRSISEN-SLVAMDFSG---QKSRVIENPTEALSVAV-EEGLAW 40

Db

143 VKRVKESPSIVGRDLAGLVSPKEVIVENPEGDFSVVVLDSGVKW 186

RESULT 20

MPPB\_NEUCR

ID MPPB\_NEUCR STANDARD; PRT; 476 AA.  
AC P11913;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Mitochondrial processing peptidase beta subunit, mitochondrial  
DE precursor (EC 3.4.24.64) (Beta-MPP) (Ubiquinol-cytochrome C reductase  
DE complex core protein I) (EC 1.10.2.2).  
GN PEP.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-34.  
RC STRAIN=74-OR23-1A / FGSC 987;  
RX MEDLINE=88223372; PubMed=2967109;  
RA Hawlitschek G., Schneider H., Schmidt B., Tropschug M.,  
RA Hartl F.-U., Neupert W.;  
RT "Mitochondrial protein import: identification of processing peptidase  
RT and of PEP, a processing enhancing protein.";  
RL Cell 53:795-806(1988).  
RN [2]  
RP IDENTITY WITH CYTOCHROME C REDUCTASE CORE PROTEIN I.  
RX MEDLINE=89238559; PubMed=2524007;  
RA Schulte U., Arretz M., Schneider H., Tropschug M., Wachter E.,  
RA Neupert W., Weiss H.;  
RT "A family of mitochondrial proteins involved in bioenergetics and  
RT biogenesis.";  
RL Nature 339:147-149(1989).  
CC -!- FUNCTION: Cleaves presequences (transit peptides) from  
CC mitochondrial protein precursors.  
CC -!- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C  
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH  
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY  
CC MEDIATE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.  
CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from  
CC precursor proteins imported into the mitochondrion, typically with  
CC Arg in position P2.  
CC -!- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2  
CC ferrocytochrome c.  
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY.  
CC -!- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M20928; AAA33606.1; -.  
DR PIR; A29881; A29881.  
DR MEROPS; M16.003; -.  
DR InterPro; IPR001431; Peptidase\_M16.  
DR Pfam; PF00675; Peptidase\_M16; 1.  
DR Pfam; PF05193; Peptidase\_M16\_C; 1.  
DR PROSITE; PS00143; INSULINASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Mitochondrion; Transit peptide;  
KW Oxidoreductase; Electron transport; Respiratory chain.  
FT TRANSIT 1 28 MITOCHONDRION.  
FT CHAIN 29 476 MITOCHONDRIAL PROCESSING PEPTIDASE  
FT BETA SUBUNIT.  
FT DOMAIN 150 178 ASP/GLU-RICH (ACIDIC).  
FT METAL 84 84 ZINC (BY SIMILARITY).  
FT ACT\_SITE 87 87 BY SIMILARITY.  
FT METAL 88 88 ZINC (BY SIMILARITY).  
FT METAL 164 164 ZINC (BY SIMILARITY).  
SQ SEQUENCE 476 AA; 52556 MW; BF3905A20D3945E4 CRC64;

Query Match 24.3%; Score 51.5; DB 1; Length 476;  
Best Local Similarity 35.7%; Pred. No. 36;  
Matches 15; Conservative 9; Mismatches 15; Indels 3; Gaps 2;

QY 1 PMRSISENSLVAMDFSGQKSRVIEN--PTEALSVAVEEGLAW 40  
|: || | || | : :: || ::|| ||::|  
Db 251 PVSSASILSKKKPDFIGSDIRIRDDTIPTANIAIAV-EGVSW 291

#### RESULT 21

##### RIR1\_AQUAE

ID RIR1\_AQUAE STANDARD; PRT; 801 AA.  
AC O66503;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ribonucleoside-diphosphate reductase alpha chain (EC 1.17.4.1)  
DE (Ribonucleotide reductase).  
GN NRDA OR AQ\_094.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus.";  
RL Nature 392:353-358(1998).  
CC -!- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM  
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY  
CC FOR DNA SYNTHESIS (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized  
 CC thioredoxin + H(2)O = ribonucleoside diphosphate + reduced  
 CC thioredoxin.  
 CC -!- PATHWAY: DNA replication pathway; first step.  
 CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS  
 CC (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE  
 CC LARGE CHAIN FAMILY.  
 CC

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE000673; AAC06460.1; -.  
 DR PIR; D70309; D70309.  
 DR HSSP; P00452; 2R1R.  
 DR InterPro; IPR005144; ATP.  
 DR InterPro; IPR000788; Ribonucleo\_red.  
 DR Pfam; PF03477; ATP-cone; 1.  
 DR Pfam; PF00317; ribonuc\_red\_lg; 1.  
 DR Pfam; PF02867; ribonuc\_red\_lgC; 1.  
 DR PRINTS; PR01183; RIBORDTASEM1.  
 DR PROSITE; PS00089; RIBORED\_LARGE; 1.  
 KW Oxidoreductase; DNA replication; Complete proteome.  
 FT ACT\_SITE 235 235 BY SIMILARITY.  
 FT ACT\_SITE 485 485 BY SIMILARITY.  
 FT ACT\_SITE 521 521 BY SIMILARITY.  
 FT SITE 796 796 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN  
 FT (BY SIMILARITY).  
 FT SITE 799 799 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 801 AA; 92913 MW; FF728EDC7D97C396 CRC64;

Query Match 24.3%; Score 51.5; DB 1; Length 801;  
 Best Local Similarity 42.9%; Pred. No. 63;  
 Matches 15; Conservative 5; Mismatches 6; Indels 9; Gaps 2;

Qy 18 QKSRVIENPTE-----ALSVAV----EEGLAWRKK 43  
 :: |||| | | |::|| | | | |  
 Db 171 EEGRVIELPQEMYMLIAMTLAVPEKPEERLKWAKK 205

# RESULT 22

## OPHL\_HUMAN

ID OPHL\_HUMAN STANDARD; PRT; 814 AA.  
 AC Q9UNA1; O75117; Q9BYS6; Q9BYS7; Q9UJ00;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Oligophrenin-1 like protein (GTPase regulator associated with focal  
 DE adhesion kinase).  
 GN OPHN1L OR GRAF OR KIAA0621.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Heart, Liver, and Placenta;  
 RA Xia J.H., Tang X.X., Yu K.P., Pan Q., Dai H.P.;  
 RT "Molecular cloning of human oligophrenin-1 like (OPHN1L) gene,  
 RT complete CDS.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2), DISEASE, AND VARIANT LEUKEMIA SER-417.  
 RX MEDLINE=20381355; PubMed=10908648;  
 RA Borkhardt A., Bojesen S., Haas O.A., Fuchs U., Bartelheimer D.,  
 RA Loncarevic I.F., Bohle R.M., Harbott J., Repp R., Jaeger U.,  
 RA Viehmann S., Henn T., Korth P., Scharr D., Lampert F.;  
 RT "The human GRAF gene is fused to MLL in a unique t(5;11)(q31;q23) and  
 RT both alleles are disrupted in three cases of myelodysplastic  
 RT syndrome/acute myeloid leukemia with a deletion 5q.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9168-9173(2000).  
 RN [3]  
 RP SEQUENCE OF 53-785 FROM N.A. (ISOFORMS 1 AND 2).  
 RA Bojesen S.E., Link C., Borkhardt A.;  
 RT "Genomic structure of the human GRAF gene.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE OF 62-814 FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 CC -!- FUNCTION: GTPase activating protein for RhoA.  
 CC -!- SUBUNIT: Binds to the C-terminal of pp125(FAK).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9UNA1-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9UNA1-2; Sequence=VSP\_001659;  
 CC -!- DISEASE: A form of juvenile myelomonocytic leukemia is  
 CC characterized by a chromosomal translocation t(5;11)(q31;q23) that  
 CC involves OPHN1L and MLL.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- SIMILARITY: Contains 1 Rho-GAP domain.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
 CC

WWW="<http://www.infobiogen.fr/services/chromcancer/Genes/GRAFID291.html>".

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF141884; AAD39482.1; -.  
DR EMBL; Y10388; CAA71414.2; -.  
DR EMBL; AJ309466; CAC29145.2; -.  
DR EMBL; AJ309467; CAC29145.2; JOINED.  
DR EMBL; AJ309468; CAC29145.2; JOINED.  
DR EMBL; AJ309469; CAC29145.2; JOINED.  
DR EMBL; AJ309470; CAC29145.2; JOINED.  
DR EMBL; AJ309471; CAC29145.2; JOINED.  
DR EMBL; AJ309472; CAC29145.2; JOINED.  
DR EMBL; AJ309473; CAC29145.2; JOINED.  
DR EMBL; AJ309474; CAC29145.2; JOINED.  
DR EMBL; AJ309475; CAC29145.2; JOINED.  
DR EMBL; AJ309476; CAC29145.2; JOINED.  
DR EMBL; AJ309477; CAC29145.2; JOINED.  
DR EMBL; AJ309478; CAC29145.2; JOINED.  
DR EMBL; AJ309479; CAC29145.2; JOINED.  
DR EMBL; AJ309480; CAC29145.2; JOINED.  
DR EMBL; AJ309481; CAC29145.2; JOINED.  
DR EMBL; AJ309482; CAC29145.2; JOINED.  
DR EMBL; AJ309483; CAC29145.2; JOINED.  
DR EMBL; AJ309484; CAC29145.2; JOINED.  
DR EMBL; AJ309485; CAC29145.2; JOINED.  
DR EMBL; AJ309486; CAC29145.2; JOINED.  
DR EMBL; AJ309487; CAC29145.2; JOINED.  
DR EMBL; AJ309466; CAC29146.2; -.  
DR EMBL; AJ309467; CAC29146.2; JOINED.  
DR EMBL; AJ309468; CAC29146.2; JOINED.  
DR EMBL; AJ309469; CAC29146.2; JOINED.  
DR EMBL; AJ309470; CAC29146.2; JOINED.  
DR EMBL; AJ309471; CAC29146.2; JOINED.  
DR EMBL; AJ309472; CAC29146.2; JOINED.  
DR EMBL; AJ309473; CAC29146.2; JOINED.  
DR EMBL; AJ309474; CAC29146.2; JOINED.  
DR EMBL; AJ309475; CAC29146.2; JOINED.  
DR EMBL; AJ309476; CAC29146.2; JOINED.  
DR EMBL; AJ309477; CAC29146.2; JOINED.  
DR EMBL; AJ309478; CAC29146.2; JOINED.  
DR EMBL; AJ309479; CAC29146.2; JOINED.  
DR EMBL; AJ309480; CAC29146.2; JOINED.  
DR EMBL; AJ309481; CAC29146.2; JOINED.  
DR EMBL; AJ309482; CAC29146.2; JOINED.  
DR EMBL; AJ309483; CAC29146.2; JOINED.  
DR EMBL; AJ309484; CAC29146.2; JOINED.  
DR EMBL; AJ309485; CAC29146.2; JOINED.  
DR EMBL; AJ309487; CAC29146.2; JOINED.  
DR EMBL; AB014521; BAA31596.1; -.  
DR PIR; F59430; F59430.  
DR MIM; 605370; -.  
DR HSSP; P19174; 2HSP.  
DR GO; GO:0005100; F:Rho GTPase activator activity; NAS.  
DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; NAS.  
DR GO; GO:0007399; P:neurogenesis; NAS.  
DR InterPro; IPR001849; PH.

DR InterPro; IPR000198; RhoGAP.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00620; RhoGAP; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00324; RhoGAP; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50238; RHO GAP; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW GTPase activation; SH3 domain; Alternative splicing;  
 KW Disease mutation; Chromosomal translocation; Proto-oncogene.  
 FT DOMAIN 265 369 PH.  
 FT DOMAIN 383 568 RHO-GAP.  
 FT DOMAIN 756 814 SH3.  
 FT DOMAIN 584 701 SER-RICH.  
 FT VARSPLIC 700 754 Missing (in isoform 2).  
 FT /FTid=VSP\_001659.  
 FT VARIANT 417 417 N -> S (IN LEUKEMIA).  
 FT /FTid=VAR\_013623.  
 FT CONFLICT 355 355 E -> G (IN REF. 2 AND 3).  
 SQ SEQUENCE 814 AA; 92234 MW; 5C81DBDECB32B18A CRC64;

Query Match 24.3%; Score 51.5; DB 1; Length 814;  
 Best Local Similarity 31.7%; Pred. No. 64;  
 Matches 13; Conservative 11; Mismatches 14; Indels 3; Gaps 1;

QY 3 RSISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWRKK 43  
 ||:|:|: : :|:|:|:|:|:|:|:|:|:|:  
 Db 85 RSLQEFATVLRNLEDERIRMIENASEVLITPLEK---FRKE 122

# RESULT 23

## GLPF\_STRPN

ID GLPF\_STRPN STANDARD; PRT; 234 AA.  
 AC P52281;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glycerol uptake facilitator protein.  
 GN GLPF OR SP2184 OR SPR1988.  
 OS Streptococcus pneumoniae, and  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313, 171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=P13;  
 RX MEDLINE=96015435; PubMed=7565084;  
 RA Saluja S.K., Weiser J.N.;  
 RT "The genetic basis of colony opacity in Streptococcus pneumoniae:  
 RT evidence for the effect of box elements on the frequency of  
 RT phenotypic variation."  
 RL Mol. Microbiol. 16:215-227(1995).



RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC BAA-334 / TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of *Streptococcus*  
 RT *pneumoniae*.";   
 RL Science 293:498-506(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC BAA-255 / R6;  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczyk L.C., Burgett S.,  
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
 RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,  
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RT "Genome of the bacterium *Streptococcus pneumoniae* strain R6.";   
 RL J. Bacteriol. 183:5709-5717(2001).  
 CC -!- FUNCTION: GLYCEROL ENTERS THE CELL VIA THE GLYCEROL DIFFUSION  
 CC FACILITATOR PROTEIN. THIS MEMBRANE PROTEIN FACILITATES THE  
 CC MOVEMENT OF GLYCEROL ACROSS THE CYTOPLASMIC MEMBRANE (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U12567; AAA91618.1; -.  
 DR EMBL; AE007506; AAK76235.1; -.  
 DR EMBL; AE008563; AAL00790.1; -.  
 DR PIR; A99520; A99520.  
 DR PIR; B95255; B95255.  
 DR PIR; S67937; S67937.  
 DR HSSP; P11244; 1FX8.  
 DR TIGR; SP2184; -.  
 DR InterPro; IPR000425; MIP\_family.  
 DR Pfam; PF00230; MIP; 1.  
 DR PRINTS; PR00783; MINTRINSICP.  
 DR ProDom; PD000295; MIP\_family; 1.

DR TIGRFAMs; TIGR00861; MIP; 1.  
 DR PROSITE; PS00221; MIP; 1.  
 KW Glycerol metabolism; Transport; Transmembrane; Complete proteome.  
 FT TRANSMEM 9 29 POTENTIAL.  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT TRANSMEM 61 81 POTENTIAL.  
 FT TRANSMEM 83 103 POTENTIAL.  
 FT TRANSMEM 135 155 POTENTIAL.  
 FT TRANSMEM 159 179 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT CONFLICT 44 45 GW -> V (IN REF. 1).  
 FT CONFLICT 63 63 H -> Y (IN REF. 2).  
 SQ SEQUENCE 234 AA; 24345 MW; 497670A3A6336065 CRC64;

Query Match 24.1%; Score 51; DB 1; Length 234;  
 Best Local Similarity 40.0%; Pred. No. 19;  
 Matches 12; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 11 VAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
 ||: ||: | || : ||: || |  
 Db 51 VAVFVSGKLSPAHLNPAVTIGVALKGGLPW 80

#### RESULT 24

##### G3P1\_BACSU

ID G3P1\_BACSU STANDARD; PRT; 334 AA.  
 AC P09124;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glyceraldehyde 3-phosphate dehydrogenase 1 (EC 1.2.1.12) (GAPDH) (NAD-  
 DE dependent glyceraldehyde-3-phosphate dehydrogenase).  
 GN GAPA OR GAP.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BD170;  
 RX MEDLINE=89160255; PubMed=2493629;  
 RA Viaene A., Dhaese P.;  
 RT "Sequence of the glyceraldehyde-3-phosphate dehydrogenase gene from  
 RT Bacillus subtilis."  
 RL Nucleic Acids Res. 17:1251-1251(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,

RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE OF 1-30.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=96345629; PubMed=8755892;  
 RA Graumann P., Schroeder K., Schmid R., Marahiel M.A.;  
 RT "Cold shock stress-induced proteins in *Bacillus subtilis*.";  
 RL J. Bacteriol. 178:4611-4619(1996).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20261518; PubMed=10799476;  
 RA Fillinger S., Boschi-Muller S., Azza S., Dervyn E., Branlant G.,  
 RA Aymerich S.;  
 RT "Two glyceraldehyde-3-phosphate dehydrogenases with opposite  
 RT physiological roles in a nonphotosynthetic bacterium.";  
 RL J. Biol. Chem. 275:14031-14037(2000).  
 CC -!- FUNCTION: More active in catabolism.  
 CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +  
 CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.  
 CC -!- PATHWAY: Second phase of glycolysis; first step.  
 CC -!- SUBUNIT: Homotetramer.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate  
 CC dehydrogenase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13011; CAA31434.1; -.  
 DR EMBL; Z99121; CAB15399.1; -.  
 DR PIR; S02754; DEBSG.

DR HSSP; P00362; 1GD1.  
 DR SubtiList; BG10827; gapA.  
 DR InterPro; IPR000173; GAP\_dhydrogenase.  
 DR InterPro; IPR006424; GAPDH-I.  
 DR Pfam; PF00044; gpdh; 1.  
 DR Pfam; PF02800; gpdh\_C; 1.  
 DR PRINTS; PR00078; G3PDHDRGNASE.  
 DR TIGRFAMs; TIGR01534; GAPDH-I; 1.  
 DR PROSITE; PS00071; GAPDH; 1.  
 KW Glycolysis; Oxidoreductase; NAD; Multigene family; Complete proteome.  
 FT INIT\_MET 0 0  
 FT BINDING 151 151 GLYCERALDEHYDE 3-PHOSPHATE.  
 FT ACT\_SITE 178 178 ACTIVATES THIOL GROUP DURING CATALYSIS.  
 SQ SEQUENCE 334 AA; 35701 MW; 1283D3E6CF5095EC CRC64;

Query Match 24.1%; Score 51; DB 1; Length 334;  
 Best Local Similarity 35.0%; Pred. No. 28;  
 Matches 14; Conservative 8; Mismatches 10; Indels 8; Gaps 2;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEG-----LAW 40  
 || ||: ||: ||: : :||| || :||  
 Db 275 SEEPLVSGDYNGNKN---SSTIDALSTMVMEGSMVKVISW 311

#### RESULT 25

##### TPS1\_PICAN

ID TPS1\_PICAN STANDARD; PRT; 475 AA.  
 AC 094213;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Alpha,alpha-trehalose-phosphate synthase [UDP-forming] (EC 2.4.1.15)  
 DE (Trehalose-6-phosphate synthase) (UDP-glucose-glucosephosphate  
 DE glucosyltransferase).  
 GN TPS1.  
 OS Pichia angusta (Yeast) (Hansenula polymorpha).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Pichia.  
 OX NCBI\_TaxID=4905;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99350434; PubMed=10419968;  
 RA Reinders A., Romano I., Wiemken A., De Virgilio C.;  
 RT "The thermophilic yeast hansenula polymorpha does not require  
 RT trehalose synthesis for growth at high temperatures but does for  
 RT normal acquisition of thermotolerance.";  
 RL J. Bacteriol. 181:4665-4668(1999).  
 CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-glucose 6-phosphate = UDP +  
 CC alpha,alpha-trehalose 6-phosphate.  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 20.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AJ010725; CAB38058.1; -.

DR InterPro; IPR001830; Glyco\_trans\_20.

DR Pfam; PF00982; Glyco\_transf\_20; 1.

KW Transferase; Glycosyltransferase; Glycolysis.

SQ SEQUENCE 475 AA; 54407 MW; 14F1A07AE88E12AB CRC64;

Query Match 24.1%; Score 51; DB 1; Length 475;

Best Local Similarity 39.5%; Pred. No. 41;

Matches 15; Conservative 6; Mismatches 11; Indels 6; Gaps 2;

QY 7 ENSLVAMDFSGQ---KSRVIENP--TEALSVAVEEGL 38

: ||| :|:| :: || || || |: |||

Db 401 KGSLVLSEFAGAAQSLNGALVVNPWNTEELSEAIYEGL 438

#### RESULT 26

##### BM3R\_BACME

ID BM3R\_BACME STANDARD; PRT; 192 AA.

AC P43506;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Transcriptional repressor Bm3R1.

GN BM3R1.

OS Bacillus megaterium.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1404;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92184811; PubMed=1544926;

RA Shaw G.C., Fulco A.J.;

RT "Barbiturate-mediated regulation of expression of the cytochrome

RT P450BM-3 gene of Bacillus megaterium by Bm3R1 protein.";

RL J. Biol. Chem. 267:5515-5526(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=89291834; PubMed=2544578;

RA Ruettinger R.T., Wen L.P., Fulco A.J.;

RT "Coding nucleotide, 5' regulatory, and deduced amino acid sequences

RT of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450

RT reductase from Bacillus megaterium.";

RL J. Biol. Chem. 264:10987-10995(1989).

RN [3]

RP CHARACTERIZATION.

RX MEDLINE=93155125; PubMed=8428974;

RA Shaw G.C., Fulco A.J.;

RT "Inhibition by barbiturates of the binding of Bm3R1 repressor to its

RT operator site on the barbiturate-inducible cytochrome P450BM-3 gene

RT of Bacillus megaterium.";

RL J. Biol. Chem. 268:2997-3004(1993).

CC -!- FUNCTION: NEGATIVELY CONTROLS THE EXPRESSION OF THE CYTOCHROME

CC P450BM-3 GENE AT THE TRANSCRIPTIONAL LEVEL.

CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; S87512; AAB21757.1; -.  
DR EMBL; J04832; AAA87601.1; -.  
DR PIR; A42116; A42116.  
DR InterPro; IPR001647; HTH\_Tetr.  
DR Pfam; PF00440; tetr; 1.  
DR PRINTS; PR00455; HTH\_TETR.  
DR PROSITE; PS01081; HTH\_TETR\_FAMILY; 1.  
KW Transcription regulation; Repressor; DNA-binding.  
FT DNA\_BIND 28 47 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 192 AA; 21886 MW; 766AC6DD34944748 CRC64;

Query Match 23.8%; Score 50.5; DB 1; Length 192;  
Best Local Similarity 35.6%; Pred. No. 18;  
Matches 16; Conservative 9; Mismatches 13; Indels 7; Gaps 3;

Qy 2 MRSISENSLVAMDFSG--QKSRVIENP----TEALSVAVEEGLAW 40  
:|::||:|:| : :||| |:| ||| | |  
Db 142 IRNLPENALIAILFGSPMEVYEMIENDYLSLTDELLTGVEESL-W 185

#### RESULT 27

##### VG13\_BPML5

ID VG13\_BPML5 STANDARD; PRT; 593 AA.  
AC Q05219;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Gene 13 protein (GP13).  
GN 13.  
OS Mycobacteriophage L5.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC L5-like viruses.  
OX NCBI\_TaxID=31757;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93211282; PubMed=8459766;  
RA Hatfull G.F., Sarkis G.J.;  
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:  
RT a phage system for mycobacterial genetics."  
RL Mol. Microbiol. 7:395-405(1993).  
CC -!- SIMILARITY: BELONGS TO THE PHAGE TERMINASE FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC -----  
DR EMBL; Z18946; CAA79389.1; -.  
DR PIR; S30958; S30958.  
DR InterPro; IPR005021; Phage\_termin.  
DR Pfam; PF03354; Phage\_terminase; 1.  
SQ SEQUENCE 593 AA; 66218 MW; EF9F3BC7B240CC66 CRC64;

Query Match 23.8%; Score 50.5; DB 1; Length 593;  
Best Local Similarity 42.9%; Pred. No. 62;  
Matches 15; Conservative 3; Mismatches 16; Indels 1; Gaps 1;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
| |: || | ||: | : | | || || |  
Db 492 SPNNPVAFDMRGQQKRFAFD-CERLEDAVLEGEVW 525

RESULT 28

VG13\_BPMD2

ID VG13\_BPMD2 STANDARD; PRT; 595 AA.  
AC O64206;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Gene 13 protein (GP13).  
GN 13.  
OS Mycobacteriophage D29.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=28369;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98300335; PubMed=9636706;  
RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;  
RT "Genome structure of mycobacteriophage D29: implications for phage  
RT evolution.";  
RL J. Mol. Biol. 279:143-164(1998).  
CC -!- SIMILARITY: BELONGS TO THE PHAGE TERMINASE FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF022214; AAC18453.1; -.  
DR PIR; B72801; B72801.  
DR InterPro; IPR005021; Phage\_termin.  
DR Pfam; PF03354; Phage\_terminase; 1.  
SQ SEQUENCE 595 AA; 66397 MW; AFD123ED5371E263 CRC64;

Query Match 23.8%; Score 50.5; DB 1; Length 595;  
Best Local Similarity 42.9%; Pred. No. 62;  
Matches 15; Conservative 3; Mismatches 16; Indels 1; Gaps 1;

QY 6 SENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40  
| |: || | ||: | : | | || || |

## RESULT 29

## ABG1\_HUMAN

ID ABG1\_HUMAN STANDARD; PRT; 678 AA.  
AC P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;  
AC Q9BXL3; Q9BXL4;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)  
DE (ATP-binding cassette transporter 8).  
GN ABCG1 OR ABC8 OR WHT1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).  
RC TISSUE=Retina;  
RX MEDLINE=96256850; PubMed=8659545;  
RA Chen H.M., Rossier C., Lalioti M.D., Lynn A., Chakravarti A.,  
RA Perrin G., Antonarakis S.E.;  
RT "Cloning of the cDNA for a human homologue of the Drosophila white  
RT gene and mapping to chromosome 21q22.3.";  
RL Am. J. Hum. Genet. 59:66-75(1996).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20289799; PubMed=10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,  
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,  
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehrach H., Reinhardt R., Yaspo M.-L.;  
RT "The DNA sequence of human chromosome 21.";  
RL Nature 405:311-319(2000).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20408883; PubMed=10950923;  
RA Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,  
RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,  
RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,  
RA Antonarakis S.E., Bonne-Tamir B.;  
RT "Refined localization of autosomal recessive nonsyndromic deafness  
RT DFNB10 locus using 34 novel microsatellite markers, genomic  
RT structure, and exclusion of six known genes in the region.";  
RL Genomics 68:22-29(2000).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 1).



RX MEDLINE=21192304; PubMed=11279031;  
 RA Porsch-Oezcueruemez M., Langmann T., Heimerl S., Borsukova H.,  
 RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;  
 RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor  
 RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene  
 RT expression and a modulator of cellular lipid efflux.";  
 RL J. Biol. Chem. 276:12427-12433(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).  
 RX MEDLINE=21092576; PubMed=11162488;  
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,  
 RA Assmann G., Cullen P.;  
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";  
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).  
 RN [6]  
 RP SEQUENCE OF 33-678 FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97186700; PubMed=9034316;  
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,  
 RA Goldenson D., Arciniegas S., Son D., Wu R.;  
 RT "Isolation and characterization of a mammalian homolog of the  
 RT Drosophila white gene.";  
 RL Gene 185:77-85(1997).  
 RN [7]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20261604; PubMed=10799558;  
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,  
 RA Mangelsdorf D.J., Edwards P.A.;  
 RT "Human white/murine ABC8 mRNA levels are highly induced in  
 RT lipid-loaded macrophages. A transcriptional role for specific  
 RT oxysterols.";  
 RL J. Biol. Chem. 275:14700-14707(2000).  
 RN [8]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20105556; PubMed=10639163;  
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,  
 RA Porsch-Oezcueruemez M., Liebisch G., Kapinsky M., Diederich W.,  
 RA Drobnik W., Dean M., Allikmets R., Schmitz G.;  
 RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a  
 RT regulator of macrophage cholesterol and phospholipid transport.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is  
 CC an active component of the macrophage lipid export complex. Could  
 CC also be involved in intracellular lipid transport processes. The  
 CC role in cellular lipid homeostasis may not be limited to  
 CC macrophages.  
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners  
 CC of the ABCG subfamily.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly  
 CC localized in the intracellular compartments mainly associated with  
 CC the endoplasmic reticulum (ER) and Golgi membranes.

CC    -!- ALTERNATIVE PRODUCTS:  
 CC        Event=Alternative splicing; Named isoforms=7;  
 CC        Comment=Additional isoforms seem to exist;  
 CC        Name=1;  
 CC        IsoId=P45844-1; Sequence=Displayed;  
 CC        Name=2; Synonyms=J;  
 CC        IsoId=P45844-2; Sequence=VSP\_000047, VSP\_000051;  
 CC        Name=3; Synonyms=ABDE;  
 CC        IsoId=P45844-3; Sequence=VSP\_000048, VSP\_000051;  
 CC        Name=4; Synonyms=G;  
 CC        IsoId=P45844-4; Sequence=VSP\_000051;  
 CC        Name=5; Synonyms=F;  
 CC        IsoId=P45844-5; Sequence=VSP\_000049, VSP\_000051;  
 CC        Name=6; Synonyms=HI;  
 CC        IsoId=P45844-6; Sequence=VSP\_000046, VSP\_000051;  
 CC        Name=7; Synonyms=C;  
 CC        IsoId=P45844-7; Sequence=VSP\_000050, VSP\_000051;  
 CC    -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.  
 CC    -!- INDUCTION: Strongly induced in monocyte-derived macrophages during  
 CC        cholesterol influx. Conversely, mRNA and protein expression are  
 CC        suppressed by lipid efflux. Induction is mediated by the liver X  
 CC        receptor/retinoide X receptor (LXR/RXR) pathway.  
 CC    -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)  
 CC        SUBFAMILY.

CC    -----  
 CC    This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC    the European Bioinformatics Institute. There are no restrictions on its  
 CC    use by non-profit institutions as long as its content is in no way  
 CC    modified and this statement is not removed. Usage by and for commercial  
 CC    entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC    or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC    -----

DR    EMBL; X91249; CAA62631.1; ALT\_INIT.  
 DR    EMBL; AP001746; BAA95530.1; ALT\_INIT.  
 DR    EMBL; AB038161; BAB13728.2; ALT\_INIT.  
 DR    EMBL; AJ289137; CAC00730.1; ALT\_INIT.  
 DR    EMBL; AJ289138; CAC00730.1; JOINED.  
 DR    EMBL; AJ289139; CAC00730.1; JOINED.  
 DR    EMBL; AJ289140; CAC00730.1; JOINED.  
 DR    EMBL; AJ289141; CAC00730.1; JOINED.  
 DR    EMBL; AJ289142; CAC00730.1; JOINED.  
 DR    EMBL; AJ289143; CAC00730.1; JOINED.  
 DR    EMBL; AJ289144; CAC00730.1; JOINED.  
 DR    EMBL; AJ289145; CAC00730.1; JOINED.  
 DR    EMBL; AJ289146; CAC00730.1; JOINED.  
 DR    EMBL; AJ289147; CAC00730.1; JOINED.  
 DR    EMBL; AJ289148; CAC00730.1; JOINED.  
 DR    EMBL; AJ289149; CAC00730.1; JOINED.  
 DR    EMBL; AJ289150; CAC00730.1; JOINED.  
 DR    EMBL; AJ289151; CAC00730.1; JOINED.  
 DR    EMBL; AF323658; AAK28836.1; -.  
 DR    EMBL; AF323644; AAK28836.1; JOINED.  
 DR    EMBL; AF323645; AAK28836.1; JOINED.  
 DR    EMBL; AF323646; AAK28836.1; JOINED.  
 DR    EMBL; AF323647; AAK28836.1; JOINED.  
 DR    EMBL; AF323648; AAK28836.1; JOINED.

DR EMBL; AF323649; AAK28836.1; JOINED.  
 DR EMBL; AF323650; AAK28836.1; JOINED.  
 DR EMBL; AF323651; AAK28836.1; JOINED.  
 DR EMBL; AF323652; AAK28836.1; JOINED.  
 DR EMBL; AF323653; AAK28836.1; JOINED.  
 DR EMBL; AF323654; AAK28836.1; JOINED.  
 DR EMBL; AF323655; AAK28836.1; JOINED.  
 DR EMBL; AF323656; AAK28836.1; JOINED.  
 DR EMBL; AF323657; AAK28836.1; JOINED.  
 DR EMBL; AF323664; AAK28842.1; -.  
 DR EMBL; AF323658; AAK28833.1; -.  
 DR EMBL; AF323640; AAK28833.1; JOINED.  
 DR EMBL; AF323645; AAK28833.1; JOINED.  
 DR EMBL; AF323646; AAK28833.1; JOINED.  
 DR EMBL; AF323647; AAK28833.1; JOINED.  
 DR EMBL; AF323648; AAK28833.1; JOINED.  
 DR EMBL; AF323649; AAK28833.1; JOINED.  
 DR EMBL; AF323650; AAK28833.1; JOINED.  
 DR EMBL; AF323651; AAK28833.1; JOINED.  
 DR EMBL; AF323652; AAK28833.1; JOINED.  
 DR EMBL; AF323653; AAK28833.1; JOINED.  
 DR EMBL; AF323654; AAK28833.1; JOINED.  
 DR EMBL; AF323655; AAK28833.1; JOINED.  
 DR EMBL; AF323656; AAK28833.1; JOINED.  
 DR EMBL; AF323657; AAK28833.1; JOINED.  
 DR EMBL; AF323660; AAK28838.1; -.  
 DR EMBL; AF323663; AAK28841.1; ALT\_INIT.  
 DR EMBL; AF323658; AAK28835.1; -.  
 DR EMBL; AF323642; AAK28835.1; JOINED.  
 DR EMBL; AF323645; AAK28835.1; JOINED.  
 DR EMBL; AF323646; AAK28835.1; JOINED.  
 DR EMBL; AF323647; AAK28835.1; JOINED.  
 DR EMBL; AF323648; AAK28835.1; JOINED.  
 DR EMBL; AF323649; AAK28835.1; JOINED.

Query Match 23.8%; Score 50.5; DB 1; Length 678;  
 Best Local Similarity 34.9%; Pred. No. 71;  
 Matches 15; Conservative 4; Mismatches 23; Indels 1; Gaps 1;

QY 2 MRSISENSLVAMDFSGQKSRVIEN-PTEALSVAVEEGLAWRKK 43  
 : : : | | | | | : | | | |  
 Db 53 LKKVDNNLTEAQRFS LPRRAAVNIEFRDLSYSVPEGPWWRKK 95

# RESULT 30

## YIS2\_YEAST

ID YIS2\_YEAST STANDARD; PRT; 993 AA.  
 AC P40562;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Putative ATP-dependent RNA helicase YIR002C.  
 GN YIR002C OR YIB2C.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RX MEDLINE=95282515; PubMed=7762303;  
 RA Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,  
 RA Schwager C., Zimmermann J., Sander C., Ansorge W.;  
 RT "Nucleotide sequence and analysis of the centromeric region of yeast  
 RT chromosome IX.";  
 RL Yeast 11:61-78(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX PubMed=9169870;  
 RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,  
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,  
 RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,  
 RA Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,  
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX.";  
 RL Nature 387:84-87(1997).  
 CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X79743; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; Z38062; CAA86204.1; -.  
 DR PIR; S48436; S48436.  
 DR SGD; S0001441; MPH1.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0003724; F:RNA helicase activity; IMP.  
 DR GO; GO:0006281; P:DNA repair; IMP.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002464; DEAH\_box.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 DR SMART; SM00490; HELICc; 1.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; FALSE\_NEG.  
 KW Hypothetical protein; ATP-binding; RNA-binding; Helicase.  
 FT NP\_BIND 107 114 ATP (POTENTIAL).  
 FT SITE 209 212 DEAH BOX.  
 SQ SEQUENCE 993 AA; 114057 MW; 474DDC99C543171F CRC64;

Query Match 23.8%; Score 50.5; DB 1; Length 993;  
 Best Local Similarity 35.3%; Pred. No. 1.1e+02;  
 Matches 12; Conservative 5; Mismatches 8; Indels 9; Gaps 2;

QY 8 NSLVAMDFSGQKSRVIENPTEALSVAVEEGLAWR 41  
 | : || | : : ||| : || : ||

Db

325 NAFKAMQ---QSQKIIANPT-----IPEGIKWR 349

## RESULT 31

RRPL\_VSVJO

ID RRPL\_VSVJO STANDARD; PRT; 2109 AA.

AC P16379;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)  
DE (L protein).

GN L.

OS Vesicular stomatitis virus (serotype New Jersey / strain Ogden).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Rhabdoviridae; Vesiculovirus.

OX NCBI\_TaxID=11283;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90177235; PubMed=2155516;

RA Barik S., Rud E.W., Luk D., Banerjee A.K., Kang C.Y.;

RT "Nucleotide sequence analysis of the L gene of vesicular stomatitis

RT virus (New Jersey serotype): identification of conserved domains in L

RT proteins of nonsegmented negative-strand RNA viruses.";

RL Virology 175:332-337(1990).

CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE  
CC POLYMERASE. IT MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS  
CC METHYLATION OF CAPS, AND POLY(A) SYNTHESIS.CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA}(N).CC -!- SUBUNIT: THOUGHT TO FORM A TRANSCRIPTION COMPLEX WITH THE  
CC NUCLEOCAPSID (N) PROTEIN.CC -!- SIMILARITY: WITH THE L PROTEIN OF OTHER RHABDOVIRUSES AND  
CC PARAMYXOVIRUSES.CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M29788; AAA48442.1; -.

DR PIR; A46309; A46309.

DR InterPro; IPR002877; FtsJ.

DR InterPro; IPR007098; RNA\_pol\_monon.

DR InterPro; IPR001016; Viral\_RNA\_pol\_L.

DR Pfam; PF01728; FtsJ; 1.

DR Pfam; PF00946; Paramyx\_RNA\_pol; 1.

KW Transferase; RNA-directed RNA polymerase.

SQ SEQUENCE 2109 AA; 242111 MW; 724CF90ECE26CAB9 CRC64;

Query Match 23.8%; Score 50.5; DB 1; Length 2109;

Best Local Similarity 39.3%; Pred. No. 2.5e+02;

Matches 11; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

QY 16 SQQKSRVIEN-PTEALSVAVEEGLAWRK 42  
:| || :: |||: |:|: |:| |  
Db 2045 NGNKSEPFDSMVAEALTKSVDKSLSWRK 2072

RESULT 32

NADA\_HELPJ

ID NADA\_HELPJ STANDARD; PRT; 336 AA.  
AC Q9ZJN1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Quinolate synthetase A.  
GN NADA OR JHP1274.  
OS Helicobacter pylori J99 (Campylobacter pylori J99).  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
OC Helicobacteraceae; Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RA Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
RT gastric pathogen Helicobacter pylori.";  
RL Nature 397:176-180(1999).  
CC -!- FUNCTION: Catalyzes the condensation of iminoaspartate with  
CC dihydroxyacetone phosphate to form quinolate.  
CC -!- PATHWAY: NAD biosynthesis; aspartate to NaMN; second step.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE QUINOLINATE SYNTHETASE A FAMILY.  
CC SUBFAMILY 3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE001550; AAD06846.1; -.  
DR PIR; A71828; A71828.  
DR HAMAP; MF\_00569; -; 1.  
DR InterPro; IPR003473; NadA.  
DR Pfam; PF02445; NadA; 1.  
DR TIGRFAMs; TIGR00550; nadA; 1.  
KW Pyridine nucleotide biosynthesis; Complete proteome.  
SQ SEQUENCE 336 AA; 37890 MW; 0299B6A4FDD53D3E CRC64;

Query Match 23.6%; Score 50; DB 1; Length 336;  
Best Local Similarity 34.9%; Pred. No. 38;  
Matches 15; Conservative 9; Mismatches 13; Indels 6; Gaps 2;



Qy 11 VAMDFSGQKSRVIENPTEALSVAVEEGL 38  
 :|: | || |:: :|| ||::||  
 Db 102 IAVKFHGQWQVVLQKLHDSISFAVDQGL 129

RESULT 34

GYS\_CAEEL

ID GYS\_CAEEL STANDARD; PRT; 672 AA.  
 AC Q9U2D9;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative glycogen [starch] synthase (EC 2.4.1.11).  
 GN Y46G5A.31.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wallis J.M.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: TRANSFERS THE GLYCOSYL RESIDUE FROM UDPG TO THE  
 CC NONREDUCING END OF ALPHA-1,4-GLUCAN.  
 CC -!- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-alpha-D-glucosyl}(N) =  
 CC UDP + {(1,4)-alpha-D-glucosyl}(N+1).  
 CC -!- PATHWAY: Glycogen biosynthesis.  
 CC -!- SIMILARITY: BELONGS TO THE MAMMALIAN/FUNGAL GLYCOGEN SYNTHASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL110485; CAB60373.1; -.  
 DR WormPep; Y46G5A.31; CE24302.  
 KW Hypothetical protein; Glycogen biosynthesis; Transferase;  
 KW Glycosyltransferase.  
 FT BINDING 56 56 UDP-GLUCOSE (BY SIMILARITY).  
 SQ SEQUENCE 672 AA; 76458 MW; 3B3C3E9044CAC8A0 CRC64;

Query Match 23.3%; Score 49.5; DB 1; Length 672;  
 Best Local Similarity 34.3%; Pred. No. 96;  
 Matches 12; Conservative 6; Mismatches 14; Indels 3; Gaps 1;

Qy 4 SISENSLVAMDFSGQKSR---VIENPTEALSVAVE 35  
 |: |: | || | | :: | | | ::  
 Db 584 SVQELAQVMYDFCGQSRQRRIILRNSNEGLSALLD 618

RESULT 35

LYS4\_YEAST



ID LYS4\_YEAST STANDARD; PRT; 693 AA.  
 AC P49367;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Homoaconitase, mitochondrial precursor (EC 4.2.1.36) (Homoaconitate  
 DE hydratase).  
 GN LYS4 OR YDR234W OR YD9934.18.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GRF88;  
 RA Gamonet F., Lauquin J.M.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Irvin S.D., Bhattacharjee J.K.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-324 FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: RESPONSIBLE FOR THE DEHYDRATION OF CIS-HOMOACONITATE TO  
 CC HOMOISOCITRIC ACID.  
 CC -!- CATALYTIC ACTIVITY: 2-hydroxybutane-1,2,4-tricarboxylate = but-1-  
 CC ene-1,2,4-tricarboxylate + H(2)O.  
 CC -!- COFACTOR: Binds 1 4Fe-4S cluster per subunit (By similarity).  
 CC -!- PATHWAY: Lysine biosynthesis; alpha-aminoadipic acid pathway;  
 CC third step.  
 CC -!- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X93502; CAA63764.1; -.  
 DR EMBL; U46154; AAA88902.1; -.  
 DR EMBL; Z48612; CAA88513.1; -.  
 DR PIR; S61067; S61067.  
 DR SGD; S0002642; LYS4.  
 DR GO; GO:0005777; C:peroxisome; IDA.  
 DR GO; GO:0019878; P:lysine biosynthesis, aminoadipic pathway; NAS.  
 DR InterPro; IPR000573; Aconitase\_C.  
 DR InterPro; IPR001030; Aconitase\_N.  
 DR InterPro; IPR004418; Homoaconitase.  
 DR Pfam; PF00330; aconitase; 1.  
 DR Pfam; PF00694; Aconitase\_C; 1.  
 DR PRINTS; PR00415; ACONITASE.  
 DR ProDom; PD000511; Aconitase\_N; 1.

DR TIGRFAMS; TIGR00139; h\_ aconitase; 1.  
 DR PROSITE; PS00450; ACONITASE\_1; 1.  
 DR PROSITE; PS01244; ACONITASE\_2; 1.  
 KW Lysine biosynthesis; Lyase; Mitochondrion; Transit peptide;  
 KW Iron-sulfur.  
 FT TRANSIT 1 20 MITOCHONDRION (POTENTIAL).  
 FT CHAIN 21 693 HOMOACONITASE.  
 FT METAL 340 340 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 407 407 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 SQ SEQUENCE 693 AA; 75150 MW; 9342E3CF83FE3FD2 CRC64;

Query Match 23.3%; Score 49.5; DB 1; Length 693;  
 Best Local Similarity 43.3%; Pred. No. 99;  
 Matches 13; Conservative 3; Mismatches 5; Indels 9; Gaps 1;

Qy 15 FSGQKSRVIENP-----TEALSVAVE 35  
 ||| |: :||| ||| :||  
 Db 474 FSGVKTEIIENPVVEEEVNAQTEAPKQSVE 503

# RESULT 36

## Y4EB\_RHISN

ID Y4EB\_RHISN STANDARD; PRT; 104 AA.  
 AC P55425;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 11.6 kDa protein Y4EB.  
 GN Y4EB.  
 OS Rhizobium sp. (strain NGR234).  
 OG Plasmid sym pNGR234a.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
 RL Nature 387:394-401(1997).  
 CC -!- SIMILARITY: NONE OBVIOUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000070; AAB92446.1; -.  
 DR Pfam; PF05284; DUF736; 1.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 104 AA; 11580 MW; 1C371D3F016FC368 CRC64;

Query Match 23.1%; Score 49; DB 1; Length 104;  
 Best Local Similarity 39.4%; Pred. No. 14;  
 Matches 13; Conservative 7; Mismatches 5; Indels 8; Gaps 2;

QY 19 KSRV--IENPTE-----ALSVAVEEGLAWRKK 43  
 |:|: ||||: : ||| | |:|:  
 Db 28 KARIGRIENPSDKGPHFRIYAGAVELGAAWQKR 60

# RESULT 37

## GATB\_METKA

ID GATB\_METKA STANDARD; PRT; 461 AA.  
 AC Q8TWS2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B  
 DE (EC 6.3.5.-) (Asp/Glu-ADT subunit B).  
 GN GATB OR MK0960.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 CC -!- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)  
 CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-  
 CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both  
 CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction  
 CC takes place in the presence of glutamine and ATP through an  
 CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By  
 CC similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
 CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.  
 CC -!- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP  
 CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.  
 CC -!- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE GATB/GATE FAMILY. GATB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE010386; AAM02173.1; -.

DR HAMAP; MF\_00121; -, 1.  
 DR InterPro; IPR004413; GatB.  
 DR InterPro; IPR006107; GatB\_cent.  
 DR InterPro; IPR006075; GatB\_N.  
 DR InterPro; IPR003789; GatB\_Yqey.  
 DR Pfam; PF01162; GatB; 1.  
 DR Pfam; PF02934; GatB\_N; 1.  
 DR Pfam; PF02637; GatB\_Yqey; 1.  
 DR TIGRFAMs; TIGR00133; gatB; 1.  
 DR PROSITE; PS01234; GATB; 1.  
 KW Protein biosynthesis; Ligase; Complete proteome..  
 SQ SEQUENCE 461 AA; 53159 MW; 2A5FFBE0E861506A CRC64;

Query Match 23.1%; Score 49; DB 1; Length 461;  
 Best Local Similarity 33.3%; Pred. No. 74;  
 Matches 12; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

Qy 1 PMRSISENSLVAMDFSGQKSRVIENTEALSVAVEE 36  
 |: || |: : : ||: | . ||:  
 Db 385 PVEIIEENGLLKVSDEDLARVVEEVIEENPQAVED 420

# RESULT 38

## HLYB\_SERMA

ID HLYB\_SERMA STANDARD; PRT; 557 AA.  
 AC P15321;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Hemolysin activator protein precursor.  
 GN SHLB.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-24.  
 RC STRAIN=SN8, and K38;  
 RX MEDLINE=88257037; PubMed=3290200;  
 RA Poole K., Schiebel E., Braun V.;  
 RT "Molecular characterization of the hemolysin determinant of Serratia  
 RT marcescens.";  
 RL J. Bacteriol. 170:3177-3188(1988).  
 CC -!- FUNCTION: INTERACTS WITH THE CELL-BOUND HEMOLYSIN. NECESSARY FOR  
 CC THE EXTRACELLULAR SECRETION AND ACTIVATION OF THE HEMOYSIN.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- SIMILARITY: STRONG, TO P.MIRABILIS HPMB.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M19.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```

DR      EMBL; M22618; AAA50322.1; -.
DR      PIR; B28182; B28182.
DR      MEROPS; M19.UNW; -.
DR      InterPro; IPR005565; HlyB.
DR      Pfam; PF03865; HlyB; 1.
KW      Hemolysis; Outer membrane; Signal; Transmembrane.
FT      SIGNAL          1          18
FT      CHAIN           19         557      HEMOLYSIN ACTIVATOR PROTEIN.
FT      TRANSMEM        277        296      POTENTIAL.
SQ      SEQUENCE        557 AA;  61916 MW;  033D777BBF5B14B1 CRC64;

```

```
QY          5 ISENSLVAMDFSGQKSRVIENPTEALSVAVEEGLAW 40
              :| :| :| ::| :::||:|:|
Db        357 VSSPTLTIAELSASHLQILPNGVFSANLSVEQGMPW 392
```

RA Fransen M., Brees C., Baumgart E., Vanhooren J.C., Baes M.,  
 RA Mannaerts G.P., van Veldhoven P.P.;  
 RT "Identification and characterization of the putative human  
 RT peroxisomal C-terminal targeting signal import receptor.";  
 RL J. Biol. Chem. 270:7731-7736(1995).  
 CC -!- FUNCTION: BINDS TO THE C-TERMINAL PTS1-TYPE TRIPEPTIDE PEROXISOMAL  
 CC TARGETING SIGNAL (SKL-TYPE) AND PLAYS AN ESSENTIAL ROLE IN  
 CC PEROXISOMAL PROTEIN IMPORT.  
 CC -!- SUBCELLULAR LOCATION: ITS DISTRIBUTION APPEARS TO BE DYNAMIC. IT  
 CC IS PROBABLY A CYCLING RECEPTOR FOUND MAINLY IN THE CYTOPLASM AND  
 CC AS WELL ASSOCIATED TO THE PEROXISOMAL MEMBRANE THROUGH A DOCKING  
 CC FACTOR (PEX13).  
 CC -!- DISEASE: Defects in PXR1 are a cause of Zellweger syndrome-1 (ZWS-  
 CC 1), a fatal peroxisome biogenesis disorder associated with severe  
 CC abnormalities in the brain, liver and kidney. Death occurs soon  
 CC after birth. This disease is due to defective import mechanisms  
 CC for peroxisomal matrix enzymes.  
 CC -!- SIMILARITY: Contains 7 TPR repeats.  
 CC -!- SIMILARITY: STRONG, TO FUNGAL HOMOLOGS (YEAST PAS10 AND P.PASTORIS  
 CC PAS8).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U19721; AAC50103.1; -.  
 DR EMBL; Z48054; CAA88131.1; -.  
 DR EMBL; X84899; CAA59324.1; -.  
 DR PIR; A56126; A56126.  
 DR PDB; 1FCH; 06-DEC-00.  
 DR Genew; HGNC:9719; PXR1.  
 DR MIM; 600414; -.  
 DR MIM; 202370; -.  
 DR MIM; 214100; -.  
 DR GO; GO:0005778; C:peroxisomal membrane; TAS.  
 DR GO; GO:0005052; F:peroxisome targeting signal-1 receptor acti. . . ; TAS.  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 4.  
 DR SMART; SM00028; TPR; 4.  
 KW Peroxisome; Repeat; TPR repeat; Transport; Protein transport;  
 KW Disease mutation; Zellweger syndrome; 3D-structure.  
 FT REPEAT 299 331 TPR 1.  
 FT REPEAT 332 365 TPR 2.  
 FT REPEAT 366 399 TPR 3.  
 FT REPEAT 415 448 TPR 4.  
 FT REPEAT 451 484 TPR 5.  
 FT REPEAT 485 518 TPR 6.  
 FT REPEAT 519 552 TPR 7.  
 FT VARIANT 489 489 N -> K (IN NALD).  
 FT /FTId=VAR\_007543.  
 FT CONFLICT 214 214 E -> EFLKFVRQIGEGQVSLESGAGSGRAQAEQWAAEFIQ  
 FT QQ (IN REF. 3).  
 FT CONFLICT 388 388 T -> I (IN REF. 1).

SQ SEQUENCE 602 AA; 66830 MW; EA4E6FAAF5E11C55 CRC64;

Query Match 23.1%; Score 49; DB 1; Length 602;  
Best Local Similarity 34.4%; Pred. No. 99;  
Matches 11; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 11 VAMDFSGQKSRVIENPTEALSVAVEEGLAWRK 42  
| : ||: : :: | ||| : | | |  
Db 459 VLFNLSGEYDKAVDCFTAALSVRPNDYLLWNK 490

RESULT 40

PEX5\_MOUSE

ID PEX5\_MOUSE STANDARD; PRT; 639 AA.  
AC 009012;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Peroxisomal targeting signal 1 receptor (Peroxisome receptor 1)  
DE (Peroxisomal C-terminal targeting signal import receptor) (PTS1-BP)  
DE (Peroxin-5) (PTS1 receptor) (PXR1P) (PTS1R).  
GN PXR1 OR PEX5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97434211; PubMed=9288097;  
RA Baes M.I., Gressens P., Baumgart E., Carmeliet P., Casteels M.,  
RA Fransen M., Evrard P., Fahimi D., Declercq P., Collen D.,  
RA Vanveldehoven P., Mannaerts G.P.;  
RT "A mouse model for Zellweger syndrome.";  
RL Nat. Genet. 17:49-57(1997).  
CC -!- FUNCTION: BINDS TO THE C-TERMINAL PTS1-TYPE TRIPEPTIDE PEROXISOMAL  
CC TARGETING SIGNAL (SKL-TYPE) AND PLAYS AN ESSENTIAL ROLE IN  
CC PEROXISOMAL PROTEIN IMPORT.  
CC -!- SUBCELLULAR LOCATION: ITS DISTRIBUTION APPEARS TO BE DYNAMIC. IT  
CC IS PROBABLY A CYCLING RECEPTOR FOUND MAINLY IN THE CYTOPLASM AND  
CC AS WELL ASSOCIATED TO THE PEROXISOMAL MEMBRANE THROUGH A DOCKING  
CC FACTOR (PEX13).  
CC -!- SIMILARITY: Contains 7 TPR repeats.  
CC -!- SIMILARITY: STRONG, TO FUNGAL HOMOLOGS (YEAST PAS10 AND P.PASTORIS  
CC PAS8).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z97018; CAB09694.1; -.  
DR MGD; MGI:1098808; Pex5.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF00515; TPR; 4.

DR SMART; SM00028; TPR; 4.  
 KW Peroxisome; Repeat; TPR repeat; Transport; Protein transport.  
 FT REPEAT 338 370 TPR 1.  
 FT REPEAT 371 404 TPR 2.  
 FT REPEAT 405 438 TPR 3.  
 FT REPEAT 452 485 TPR 4.  
 FT REPEAT 488 521 TPR 5.  
 FT REPEAT 522 555 TPR 6.  
 FT REPEAT 556 589 TPR 7.  
 SQ SEQUENCE 639 AA; 70707 MW; 923E892D8FBB0709 CRC64;

Query Match 23.1%; Score 49; DB 1; Length 639;  
 Best Local Similarity 34.4%; Pred. No. 1.1e+02;  
 Matches 11; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

Qy 11 VAMDFSGQKSRVIENPTEALSVAVEEGLAWRK 42  
 | : ||: : :: | ||| : |||  
 Db 496 VLFNLSGEYDKAVDCFTAALSVRPNDYLMWNK 527

Search completed: January 13, 2004, 16:22:45  
 Job time : 9.11024 secs